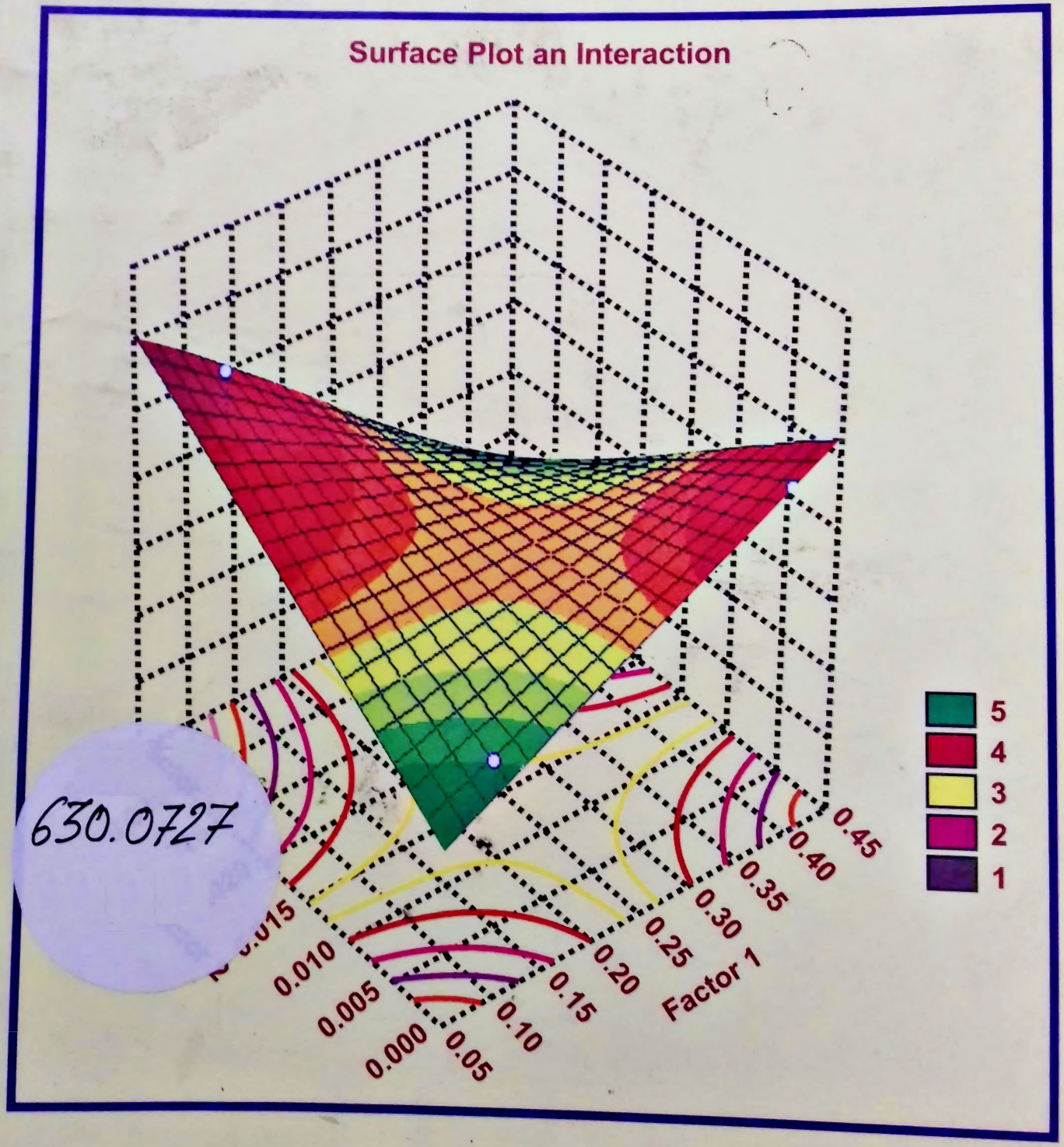
362

EXPERIMENTAL DESIGNING AND DATA ANALYSIS IN AGRICULTURE AND BIOLOGY

(Incorporates I.C.A.R. Recommended Syllabus)



Deepak Grover
Lajpat Rai

EXPERIMENTAL DESIGNING AND DATA ANALYSIS IN AGRICULTURE AND BIOLOGY
• Deepak Grover
• Lajpat Rai

630.0727

EXPERIMENTAL DESIGNING AND DATA ANALYSIS IN AGRICULTURE AND BIOLOGY

By
Deepak Grover
and
Lajpat Rai

Agrotech Publishing Academy
Udaipur - 313001

Information contained in this book has been published by Agrotech Publishing Academy and has been obtained by its authors believed to be reliable and are correct to the best of their knowledge. However, the the publisher and its authors shall in no event be liable for any errors, omissions or damage arising out of use of this information and specially disclaim any implied warranties or merchantability or fitness for any particular use. Disputes if any, are subjected to Udaipur jurisdiction only.

Published by:

Mrs. Geeta Somani
Agrotech Publishing Academy
11-A, Vinayak Complex B
Udaipur - 313001 (INDIA)
Mob: 9414169635, 9413763031

© 2010 by Authors
All rights reserved

ISBN: 987-81-8321-154-3

Typeset by:

Dayal Computers
25, Bohraganeshji,
Udaipur – 313001

Printed at:

S.S.S. Printers
New Delhi - 110002

PREFACE

It is widely believed that statistics in general and design of experiments in particular is a difficult, dull, highly specialized subject to the studied and handled by professionals who are mathematically oriented. This is a big misconception which has grown and developed over time just by 'ripple effect'. Actually, the subject is intersting, stimulating, revealing and above all near to reality as uncertainties are also taken care of while building models and drawing inferences. The need of the hour in research is to remove this misconception by putting forward practical examples, their solution in a simple way and eschewing mathematical jargon as far as possible.

This book is an attempt to correct their misconception. So that design of experiments can be introduced to be used extensively among a larger audience. Such audience includes students of agriculture, biology, statistics, research methodology, social sciences, forestry, medical sciences, environmental sciences, animal sciences, veterinary sciences, business management and engineering sciences to larger extent. In order to achieve this objective the authors have adopted an expositional style with simple concepts, tools and use with many examples from agriculture and biological sciences but the concepts and treatment remains almost same while dealing with problems from other sciences in the application of various designs discussed in this book.

The special feature of this book is its simplicity and the systematic style of presentation of the subject matter. The assumption and conditions on which various analysis and planning of the designs are based have been explained by means of simple, stepwise and straight strategy. The examples and theory have been taken from various books, manuals and reports for the benefit of students and teachers to comprehend and appreciate the actual problem faced in planning of experiments and analysis for inference. All such references have been listed at the end of book. Authours wish to acknowledge their input where ever used. Some hypothetical examples have also been considered to make the concept clear and

practically oriented. We have also tried that the implication of the results emerging from analysis be discussed with simplicity and thoroughness.

The main reason for writing this book is to reach the wider audience which has been kept out as a result of the complexity of treatment of this subjected by many authors. Also, the other objective is to bring together many concepts and designs scattered in literature for the benefit of readers. Readers from even diverse fields like economics, engineering, industries, medical etc. can find the book handy and helpful for problem encountered by them during their research and decision making. The long experience of more than thirty years of authors while offering courses on design and analysis of experiments to post-graduate students and applied workers, teachers and scientists in the university have prompted to write a book on planning and analysis of design of experiments. The book contains enough material for a course extending to a year, as different kinds, both elementary and complex design have been discussed avoiding mathematical symbols while example have been solved and presented in steps so as to understand the intricacies of analysis.

Many statistical packages are rapidly becoming available in the market as a routine resource in different institutes, colleges, universities and research centres in which a number of designs along with analysis are given which are being used sometimes unmindfully without consulting any competent statistician thus resulting in wrong inferences. Computers are unintelligent servants driven by programs. We can use any design on the same data and results will follow. It is the duty of experimenter/researchers to look into conditions and assumption and use the right design for right type of data. Hence statistical packages are to be used carefully with open eyes and mind. This book can be helpful in looking into various kinds of designs which may fit the required experiment and help the investigator for interpreting the results.

This book covers a wide range of topics arranged in nineteen chapters ranging from basic principles, analysis of variance,

transformation of data. single factor experiments, Latin square, factorial experiments, confounding, split and strip plot designs, covariance analysis, multi-locational experiments, BIBD, Laltice designs, augmented designs, block design analysis of diallel crosses, biological assays and simple way of finding expected value of mean square. Every care has been taken to put even the complex and intricate concepts in simple language for the under standing of reader may be student researcher or teacher in agricultural, engineering and biological sciences particularly.

Thanks are due to many authors of reports, investigations and books for using the examples /data for illustrating the concept and analysis of various designs for the convence and help to various readers, investigators and researchers. We are also thankful to the Director AAREM, CCS HAU and Dean, COBS&H, CCSHAU, for their help, guidance and cooperation in bringing out this book.

The author shall feel obliged for the suggestion from the readers in the improvement, modification and enhancement of contents, presentation and exposition of the book.

Hisar

18th October, 2008

Deepak Grover

Lajpat Rai

Contents

	Preface	3
1.	Basic Principles of Design of Experiments	9
2.	Analysis of Variance and Tests for Comparision Between Treatment Means	25
3.	Transformation of Data	38
4.	Single Factor Experiments	45
5.	Repeated Latin Squares, Graeco Latin Square, Cross Over and Youden Square Designs	70
6.	Sampling in Field Experiments	82
7.	Missing Plot Technique	94
8	Factorial Experiments	110
9.	Confounding	134
10.	Split and Strip Plot Design	148
11.	Covariance analysis	172
12.	Multi-Locational Experiments	180
13.	Balanced Incomplete Block Design	191
14.	Lattice Designs	202
15.	Layout and Analysis of Augmented Designs	213
16.	Block Designs	229
17.	Analysis of Diallel Crosses	248
18.	Design and Analysis for Biological Assays	274
19.	Expected Value of Mean Square in Fixed, Random & Mixed Models	288
	References	308

About the Book

The book is an attempt to put together different designs and concepts for the benefit of students, teachers, researchers and investigators in a simple and non-mathematical way. Basic concepts of planning any design along with its analysis has been presented in steps so that even the novice may be able to understand the intericasies and get the analysis done in a proper way. Book contains nineteen chapters encompassing basic principles, analysis of variance, transformation of data, single factor experiments, Latin squares, Youden square designs, missing plot techniques, factorial experiments and confounding, spell and strip plot designs, covariance analysis, BIBD, Lattice designs, Augmented designs, block designs, sampling in field experiment, multl-locational experiments, analysis of diallel crosses, design and analysis of bio-assys have also been discussed at length along with examples. Finding expected value of mean square simple method has also been given. Optimality criterion for a design is also considered. The book shall the helpful for researchers and investigators in agriculture sciences, veterionary sciences, animal sciences, biological sciences, social sciences, medical sciences, engineering and environmental sciences also.

About the Authors

Dr. Deepak Grover is working as Professor (statistics) at the CCS Haryana Agricultural University, Hisar. He has been contributing to the field of statistics since 1971 when he joined the University. He has held various posts including Head of Department, Maths and Statistics from 2003 to 2007 and taken department to newer heights by publishing manuals, research bulletins and training compediums for the benefit of teachers, researchers and students alongwith other staff members. He has taught various PG and UG courses in Statistics, Mathematics and Economics for nearly four decades. During 1996 to 1998 he taught at Addis Ababa University, Ethiopia and was consultant for staff at ESPD, United Nationas Economic Commission for Africa. He has also visited a number of UK Universities under CUSAC programme and Israel on Israel Government Fellowship. He has written/Edited five books and large number of research papers on theoretical and applied stastistics in social sciences. He is life member of more than ten scientific societies and held the post of editor, executive member and advisor. He has chaired various scientific sessions and has been invited as key note address speaker. He is founder President of Society of Applied Statistics and Computer Science. He has been on the advisory committee of more than twenty five M.Sc. and Ph.D. students of the University.

Dr. Lajpat Rai is Professor and Head, Department of Mathematics and Statistics CCS Haryana Agricultural University, Hisar. He obtained his M.Sc. degree in Agricultural Statistics from Indian Agriculture Research Institute, New Delhi and Ph.D from Haryana Agricultural University. He has published more than 40 papers in journals of repute in the area of Design of Experiments and Applied Statistics and seven manuals. Dr. Lajpat Rai is actively involved in undergraduate/postgraduate teaching of the department and has served as resource persons in many academic activities of the university. He is member of several national professional societies. He has been on the advisory committee of more than hundred M.Sc and Ph.D. Students of the University.

Chapter – 1

Basic Principles of Design of Experiments

Scientists in the biological fields who are involved in research constantly face problems associated with planning, designing and conducting experiments. Basic familiarity and understanding of statistical methods that deal with such issues of concern would be helpful in many ways. Researchers who collect data and then look for a statistical technique that would provide valid results will find that there may not be solutions to the problem and that the problem could have been avoided first by a properly designed experiment. Obviously, it is important to keep in mind that we cannot draw valid conclusions from poorly planned experiment. Second, the time and cost involved in many experiments are enormous, and a poorly designed experiment increases such costs, time and resources. The following examples point out the necessity for a good design that will yield good results in research. First, a nutrition specialist in developing country is interested in determining whether mother's milk is better than powdered milk for children under age one. The nutritionist has compared the growth of children in village A, who are all on mother's milk against the children in village B, who use powdered milk. Obviously, such a comparison ignores the health of the mothers, the sanitary conditions of the villages, and other factors that may have contributed to the difference observed without any connection to the advantages of mother's milk or the powdered milk on the children. A proper design would require that both mother's

milk and the powdered milk be alternatively used in both villages, or some other methodology to make certain that the differences observed are attributable to the type of milk consumed and not to some uncontrollable factor. Second, a crop scientist who is comparing two varieties of maize, for instance would not assign one variety to a location where such factors as sun, shade, unidirectional fertility gradient, and uneven distribution of water would either favour or handicap it over the other. If such a design were to be adopted, the researcher would have difficulty in determining whether the apparent difference in yield was due to varietal differences or resulted from such factors as sun, shade, soil fertility of the field, or the distribution of water. The two examples illustrate the type of poorly designed experiments that are to be avoided.

- Good experimental designs are product of the technical knowledge of one's field, an understanding of statistical techniques and skill in designing experiments. Any research endeavor may entail the following phases:
- Concept
- Design
- Data collection
- Analysis
- Dissemination

Statistical methodologies can be used to conduct better scientific experiments if they are incorporated into entire scientific process, i.e., from inception of the problem to experimental design, data analysis and interpretation.

When planning experiments, we must keep in mind that large uncontrolled variations are common occurrences. Experiments are generally undertaken by researchers to compare effects of several conditions on some phenomena or in discovering an unknown effect of particular process. An experiment facilities the study of such phenomena under controlled conditions. Therefore, creation of controlled conditions is the most essential characteristics of the

experimentation. How we formulate our questions and hypotheses are critical to the experimental procedure that will follow. For example, a crop scientist who plants the same variety of a crop in a field may find variations in yield that are due to periodic variations across a field or some other factors that the experimenter has no control over. The methodologies used in designing experiments will separate with confidence and accuracy a varietal difference of crops from the uncontrolled variations.

The above discussion brings some salient features of experimentation:

(i) All the extraneous variation in the data should be eliminated or controlled excepting the variation due to the treatments under study. One should not artificially provide circumstances for one treatment to show better results than others.

(ii) For a given size of the experiment, though the experiment can be done in many ways, even the best results may not turn out to be significant with some designs, while some other designs can detect the treatment differences. Thus there is an imperative need to choose the right type of design, before the commencement of the experiment.

(iii) If for some specific reason related to the nature of experiment, a particular method has to be used in experimentation, then adequate number of replications for each treatment have to be provided in order to get valid inferences.

(iv) The treatment have to be randomly allocated to the experimental units.

The Design of an experiment is a complete sequence of steps taken before the experiment to ensure that appropriate data will be obtained, data which permit an objective analysis leading to valid inferences with respect to problem at hand. The individual who formulates the design should clearly understand the objective of the investigation.

Basic Principles of Design of Experiments

Given a set of treatment which can provide information regarding the objective of an experiments, a design for the experiment defines the size and the number of experimental units, the manner in which treatments are allotted to the units and also appropriate type and grouping of the experimental units. These requirements of a design ensure validity, interpretability and accuracy of the results obtained from an analysis of the observations.

These purposes are well served by the following three principles given by Prof. Ronald A. Fisher and are Known as the Basic Principles of Design of Experiment. The three principles are:

(i) Replication

(ii) Randomization

(iii) Local control

Replication

If a treatment is allotted to r experimental units in an experiment, it is said to be replicated r times. If in a design each of the treatments is replicated r times, the design is said to have r replications. Replication is necessary to:

- Increase the accuracy of estimates of the treatment effects.
- Provide an estimate of the error variance which is a function of the differences among observations from experimental units under identical treatments.

Though, the more the number of replications the better it is, so far as precision of estimates is concerned, it cannot be increased infinitely as it increases the cost of experimentation. Moreover, due to limited availability of experimental resources too many replications cannot be taken.

The number of replication is therefore decided keeping in view the permissible expenditure and the required degree of precision. Sensitivity of statistical methods for drawing inference also depends

on the number of replications. Sometimes this criterion is used to decide the number of replications in specific experiments.

Determination of number of replications

Error variance provides a measure of precision of an experiment, the less the error variance the more precision. Once a measure of error variance is available for a set of experimental units, the number of replications needed for a desired level of sensitivity can be obtained as below.

Given a set of treatments an experimenter may not be interested to know if two treatment differ in their effects by less than a certain quantity, say, d. in other words, he wants an experiment which should be able to differentiate two treatments when they differ by d or more.

We know that the significance of the difference between two treatments is tested by t-test where

$$t = \frac{\bar{y}_i - \bar{y}_j}{\sqrt{2s^2/r}}$$

Here $\bar{y}_i$, and $\bar{y}_j$ are the arithmetic means of two treatment effects each based on r replications, s^2 is measure of error variation.

Given a difference d, between two treatment effects such that any difference greater than d should be brought out as significant by using a design with r replications, the following equation provides a solution of r.

$$t_0 = \frac{|d|}{\sqrt{2s^2/r}} \cdot$$

$$\Rightarrow \quad r = \frac{t_0^2}{d^2} x 2s^2 \qquad \ldots (1.1)$$

where t_0 is the critical value of the t-distribution at the desired level of significance, that is, the value of t at 5 or 1 percent level of significance read from the t-table. If s^2 is known or based on a very large number of observations, made available from some plot pre-

experiment investigation, then t is taken as the normal variate. If s^2 is estimated with n degree of freedom (d.f.) then t_0 corresponds to n degree of freedom.

When the number of replication is r or more as obtained above, then all differences greater than d are expected to be brought out significant by an experiment when it is conducted on as set of experimental units which has variability of the order of s^2.

For example, an experiment on wheat crop was conducted in a seed farm in Bhopal to study the effect of application of nitrogen and phosphorus on yield in a randomized block design with three replications. There were 11 treatments two of which were (i) 60kg/ha of nitrogen (ii) 120kg/ha of nitrogen. The average yield figures for these two application of the fertilizer were 1438 and 1592 kg/ha respectively and it is required that differences of the order of 150 kg/ha should be brought out significant. The error mean square (s^2) was 12134.88. Assuming that the experimental error will be of the same order in future experiments and t_0 is of the order of 2.00, which is likely as the error d.f. is likely to be more than 30 as there are 11 treatments, Substituting in (1.1); we get:

$$r = \frac{2t_0^2 s^2}{d^2} = \frac{2 \times 2^2 \times 12134.88}{150^2} = 4(\text{approx}).$$

Thus, an experiment with 4 replications is likely to bring out differences of the order of 150kg/ha as significant.

Another criterion for determining r is to take a number of replications which ensures at least 10 d.f. for the estimate of error variance in the analysis of variance of the design concerned since the sensitivity of the experiment will be very much low as the F test (which is used to draw inference in such experiment) is very much unstable below 10 d.f.

Randomization

As discussed earlier, by replication the experimenter tries to

average out as far as possible the effects due to uncontrolled factors. This brings to him the question of allocation of treatments to experimental units so that each treatment gets an equal chance of showing its worth. In the absence of the prior knowledge of the variability of the experimental material, this objectives is achieved through "randomization", a process of assigning the treatments to various experimental units in a purely chance manner. The following are the main objectives of randomization.

(i) The validity of the statistical tests of significance, e.g., the t –test for testing the significance of the difference of two means or the 'Analysis of Variance', F-test for testing the homogeneity of several means, depends on the fact that the statistic under consideration obeys some statistical distribution. Randomization provides a logical basis for that and makes it possible to draw rigorous inductive inferences by the use of statistical theories based on probability theory. This assumption of randomness is necessary since $S.E._{(\bar{x})} = \sigma/\sqrt{n}$ for random sampling only. Randomising the treatments over the experimental units is an essential safeguard against distortion of experimental results by un-anticipated influences such as rise in ambient temperature, drift in calibration of instruments and equipment, fertility of the soil or other systematic changes.

(ii) The purpose of randomness is to assure that the sources of variation, not controlled in the experiment, operate randomly so that the average effect of any group of units is zero. In other words, randomization ensures that different treatments, by the repetition of the experiment, on the average are subject to equal environmental effect. Randomization eliminates bias in any form. It equalises even factors of variation over which we have no control.

It should be noted that randomization without replication is not sufficient. It is only when randomization of treatments to various units is accompanied by an adequate number of replications then we are in a position to apply the tests of significance (t-test or F-test)

Local Control

If the experimental material, say field for agriculture experimentation, is heterogeneous and different treatments are allocated to various units (plots) at random over the entire field, the soil heterogeneity will also enter the uncontrolled factors and thus increase the experimental error. It is desirable to reduce the experimental error as far as practicable without unduly increasing the number of replications or without interfering with the statistical requirement of randomness, so that even smaller differences between treatments can be detected as significant. In addition to the principles of replication and randomization discussed earlier, the experimental error can further be reduced by making use of the fact that neighbouring areas in a field are relatively more homogeneous that those widely spread. In order to separate the soil fertility effects from the experimental error, the whole experimental area (field) is divided into homogeneous groups (blocks) row-wise or column-wise (one-way elimination of fertility gradient, c.f. Randomized Block Design) or both (elimination of fertility gradient in two perpendicular direction c.f. Latin Square Design, according the fertility gradient of the soil such that variation in each block is minimum and between the block is maximum. The treatments are then allotted at random within each block. Then the process of reducing the experimental error by dividing the relatively heterogeneous experimental area in to homogeneous blocks is known as local control. Various forms of arranging the units (plots) in to homogeneous groups (blocks) have so far been evolved and are known as experimental designs, e.g. Randomized Block Design, Latin Square Design etc.

Planning Consideration

In adding to replication, randomization and local control a few other points should be given consideration in planning field experiments. Some of the important steps are as under:

1. Objectives of the experiment
2. Scope of the experiment
3. Formulation of the Hypothesis

4. Selection of treatments
5. Selection of site and experimental material
6. Selection of plot size, shape and blocks
7. Determination of number of replications.
8. Choice of an experimental design.
9. Transformations
10. Statistical analysis
11. Drawing of conclusions.

Objective of the experiment

The preparation of a written statement of the objectives, listed in order of their priority will often result in an more concise and realistic experiment.

Scope of the experiment: The proper application of the scope of the experiment presumes a knowledge of the agricultural background of the problem and it is desirable that agricultural aspects of the experimental design should receive full consideration alongwith the statistical aspects before it is finally settled.

Formulation of the hypothesis

Keeping in view the objectives of the experiment, the researcher has to formulate certain hypothesis alongwith level of confidence which after the application of certain test of significance is to be accepted or rejected.

Selection of Treatments

The treatments to be used in an experiment are determined by the objectives. They are selected to answer the questions raised by the problem and the specific objectives. If the objective of the experiment is to evaluate new selections in a breeding program, the treatments are then new selections along with one or more standard varieties against which the new selections can be compared.

Site and experimental material

The choice of the experimental site affects the conduct of the

experiment and the interpretation of the results in several ways. It determines the population to which the results of the experiment will apply. Thus the population to which the information is wanted must be defined carefully and the site must be one that represents the population. The selected material must be homogenous and representative of the population for which results are to be made applicable.

Selection of plot size and shape and blocks

A uniformity trail is conducted to know the nature and magnitude of soil heterogeneity and to determine the optimum size and shape of plots and blocks. At the time of harvest the whole field is divided in to smaller units and the yield of these units are individually recorded. Any variation in yield is attributed to soil heterogeneity. On the basis of individual plot yield, a contour map can be prepared by joining the lines of equal fertility. There are several methods of determining the optimum plot size, some of them are (i) maximum curvature method (ii) Heterogeneity index method Smith (1938) etc.

Number of Replications

The number of replications serves the double purpose i.e. it provides an estimate of the error as well as it is the biggest force of reducing the experimental error. The deciding factors for determining the number of replications are:

(i) Degree of precision required.
(ii) Amount of variability present in the material.
(iii) Available resources.
(iv) Size and shape of experimental units.

There are many methods available for determining the optimum number of replications, however as layman the number of replications should be such that it provides atleast 10-12 degree of freedom for error.

Choice of an experimental design

There are a number of factors that affects the choice of an

experimental design, however a cardinal rule "choose the simplest experimental design that will give the required precision within limits of the available resources.

There are three main steps in choosing the design:

(i) It has to be decided whether the design is unifactor or factorial.

(ii) It has to be decided whether to group the observation to eliminate one, two or more levels of heterogeneity (causes of variations). For instance, it may be desired to eliminate simultaneously the effect of time and day of taking the observations. A Latin square design e.g.:

	Day			
Time	1	2	3	4
1	A	B	C	D
2	B	C	D	A
3	C	D	A	B
4	D	A	B	C

Illustrates one method of doing this, since each treatment is observed once each day and once at each time of the day.

(iii) If the no. of treatment combinations is too large to be adjusted in a single block, the design will be referred to as an "incomplete block design" otherwise as "complete block design".

Following table indicates the types of design that might be used for unifactor or factorial experiments in complete or incomplete blocks eliminating one or two ways of heterogeneity.

Table 1: Classification of main designs

		Unifactor	Factorial
Complete block	One grouping	Completely randomized	
		Randomized block	
	Two grouping	Latin square	
Incomplete block	One grouping	Balanced Incomplete block	Confounded design
		Partially balanced and cyclic designs	Fractional replication, split plot design
	Two grouping	Youden squares Lattice squares	Quasi-Latin squares

Transformation

Sometimes a situation arises when a straight forward analysis of variance of data is not valid. This happens when the observation refer to small whole numbers. This difficulty can be overcome by transforming the data on a suitable scale and analyzing the transformed values.

Statistical analysis

When the experiment has been completed, the data should be subjected to statistical analysis. The first step is to prepare an ANOVA Table. Once the ANOVA has been completed, the data are subjected to additional analysis to proved specific information on the questions raised by the objectives of the experiment.

Drawing of conclusion: The results are to be interpreted and put up in the form of a report.

Terminology

Experiment: An experiment is a device or a means of getting an answer to the problem under consideration. Experiment may be absolute or comparative.

Absolute experiments: Here we determine the absolute value of some characteristic like (i) obtaining the average Intelligence Quotient (IQ) of a group of people (ii) finding the correlation coefficient between two variables in a biveriate distribution etc.

Comparative experiments: Here the effect of two or more objects on some population characteristics, e.g comparison of different manures and fertilizers, different kinds of varieties of crop, different cultivation processes, different pieces of land in a field experiment, or different diets or medicines in a dietary or medical experiment respectively are studied.

Treatment

Treatment refers to controllable quantitative or qualitative factor

imposed at a certain level by the experiment. For an Agronomist several fertilizers, concentrations applied to a particular crop or a variety of crop is a treatment. Similarly, an Animal Scientist looks upon several concentrations of a drug given to animal species as a treatment. In agribusiness we may look upon impact of advertising strategy on sales of a departmental store. To an agricultural engineer, different levels of irrigation may constitute a treatment. For a plant breeder different varieties are treatments.

Experimental unit

Experimental unit may be looked upon as a small subdivision of the experimental material, which receives the treatment. For a agronomist or horticulturist it may be plot of a land or batch of seed, for an animal scientist it may be a group of pigs or sheep, for a scientist engaged in forestry research it may be different tree species occurring in an area, and for an agricultural engineer it may be manufactured item.

Experimental errors

Differences in yields arising out of experimental units treated alike are called experimental errors.

To the statistician an experiment error reflects errors of experimentations, errors of observations, errors of measurements, the variation of experimental material and the combined effects of extraneous factors which are not singled out for attention in the investigation.

Experimental error may often be reduced by the use of more homogeneous experimental material, careful stratification of available material, greater care in conducting the experiment, more efficient experimental designs, related information provided by other sources.

Yield: The measurement of a variable under study on different experimental units (e.g. plots in field experiments) are termed as yields.

Replication: Replication measures the execution of an

experiment more than once. In other words, the repetition of treatment under investigation is known as replication.

Precision: The reciprocal of the variance of the mean is termed as the precision or the amount of information of a design. Thus for an experiment repeated r times the precision is given by:

$$\frac{1}{\text{var}(x)} = \frac{r}{\sigma^2}$$ where σ^2 is the error variance per unit

Efficiency of a design

Consider two designs D_1 and D_2 with error variance per unit σ_1^2 and σ_2^2 and replications r_1 and r_2 respectively. Then the variance of difference between two treatment mean is

$$\frac{2\sigma_1^2}{r_1} \text{ and } \frac{2\sigma_2^2}{r_2} \text{ for designs } D_1 \text{ and } D_2.$$

Then efficiency of design D_1 w.r.t. D_2

$$E = \frac{\text{Amount of information in } D_1}{\text{Amount of information in } D_2} = \frac{r_1}{\sigma_1^2} \div \frac{r_2}{\sigma_2^2}$$

If $E = 1$ then both the designs are said to be equally efficient

If $E > 1$ ($E < 1$) the D_1 is said to be more (less) efficient than D_2

An efficient design has a greater ability of detecting the differences of treatment effects, so the efficiency of a design can be increased by

(i) decreasing σ^2, the error variance per unit. This is done by arranging plots in to small homogeneous blocks and

(ii) increasing r, the number of replication

Factors

Experimental variable are termed as a factor. For example, a fertilizer, a new feed ration, and a fungicide are all considered as factors.

Level of a factor

Level of a factor is the number of ways in which a factor is varied in the experiment. When a factor is varied in p ways we refer it as having p levels e.g. 2 modes of an application such as drip or sprinkle irrigation, is referred to as 2 level factors. Application of Nitrogen (N) to crop as N_0 = 0 N_1= 60 kg/ha, N_2 = 120 kg/ha is referred to as 3 level factors.

Factorial Experiment

An experiment is called factorial experiment if its treatments are combination of levels of different factors.

Uniformity trails

To conduct a uniformity trail the area is planted uniformly to a single crop. Cultural practices are conducted throughout the trial in as uniform way as possible. The trial is then partitioned into small units, or plots. These are harvested individually and the individual yields, together with the location of the plot from which they came, are recorded. These can be used to construct a yield map of the site. Such a map can be helpful in deciding the layout of the field plan. It can also be used in determining the shape and orientation of the plots and in grouping the plots into blocks of similar plots. Further information from a uniformly trial can be used to determine the optimum size of plots as well.

Agronomy trials

Agronomy trials include fertilizer studies, time, rate and density of planting, tillage study, and pest and weed control studies. Because the response to changes in the level of one factor is often conditioned by the level of other factors it is almost essential that the treatments in agronomy trials include combination of multiple levels of two or more production factors. In general the plot size for agronomy trials is large than that of variety trials. This is because of the size of machinery involved in applying the treatments and because of the

border effects introduced, in part, by the use of machinery.

Variety trials

Variety trials form the backbone of a plant breeding programme. They range from the initial studies involving small plots (single plant or single row) to replicated yield trials involving fairly large plots.

Blocking

The procedure here is to group the plots or experimental units into blocks or groups of homogeneous units. Treatments are then assigned at random to the plots within the blocks.

For example, the field under consideration for an experiment might contain two types of soil, we could select plots on each soil type and randomly assign treatments to plots within soil types. In this case soil type would be a blocking factor and the difference between soil types would be removed from the experimental error in the analysis of data. The key to successful blocking is to maximize the variation among blocks and to minimize the variation among plots within blocks.

Blocks and replications are not the same thing. A block is simply a group of plots, whereas a replication is a repetition of a treatment in an experiment.

Chapter -- 2

Analysis of Variance and Tests for Comparison between Treatment Means

The technique used for the analysis and interpretation of data collected through designing an experiment is the analysis of variance (ANOVA) which essentially consists of partitioning the total variation present in the data into meaningful components. There are mainly three kinds of variation present in the data so generated due to (i) treatments (ii) environmental (iii) error. Any variation which arises due to the factors introduced by the experimenter intentionally into the experiment will be known as variation due to treatments, while the variation in the experimental material is called environmental variation. The uncontrolled variation, which arises due to random fluctuations in the data will be labelled as random error and importantly provides a yard stick for testing the treatment and environmental variations. The partitioning of the total variation into respective components will be achieved by suitably assuming a statistical model for the observational data in terms of certain unknown parameters.

The parameters in the analysis of variance model may either be fixed constants or random variables. Depending on the nature of the parameters. Eisenhart classified the analysis of variance models into three categories as follows:

An analysis of variance model is a fixed effects model (or Eisenhart's model – I) if all the parameters are constants; random

effect model (or Eisenhart's model – II) if all the parameters excepting the general mean are random variables; or mixed effect model (or Eisenhart's model – III) if parameters other than general mean are both constants and random variables. In problems where the experimenter will be confining his attention just to the treatments applied, fixed effects model will be applied. However, to draw inferences beyond the applied treatments, random effects model will be used and in such cases the treatments and/or experimental units will be randomly chosen from the population of treatments and/or experimental units to which inferences are proposed to be made.

In fixed effect models, the main objectives are to estimate the effects, find a measure of variability among the effects of each of the factors and finally find the variability among the error effects. In random effect models, the main emphasis is on estimating the variability among the effects of the different factors. The methodology for obtaining expressions of variability is, however, mostly the same in the different models, though the methods appropriate for their testing are different.

1. Fixed Effects Model

One way classification: Let a factor A be experimented with s levels on n homogeneous experimental units, i^{th} level being used on n_i units $\left(n = \sum_{i=1}^{s} n_i\right)$. If y_{ij} is the response from the j^{th} experimental unit to which the i^{th} level is given, we can assume.

$$y_{ij} = \mu + \alpha_i + e_{ij}, \quad j = 1, 2, \ldots\ldots\ldots\ldots, n_i;\ i = 1, 2, \ldots\ldots\ldots\ldots, s$$

where μ is the general mean, α_i is the effect of i^{th} level of factor A and e_{ij}'s are random errors. Assuming that the errors are independently and normally distributed with mean zero and constant variance σ^2. The analysis of variance table partitioning the total variation into two components will be obtained as given in table 1.1.

The null hypothesis to be tested in problems of this nature is H_0 : $\alpha_2 = \alpha_2 = \ldots\ldots\ldots \alpha_s$. That is, the effects of s levels are equal against

the alternative hypothesis that not all α_i's are same. If the calculated F value in the above Anova table exceeds $F_{s-1,n-s,\alpha}$ then we reject H_0 and conclude that the effects of all the levels are not equal at α level of significance; otherwise not.

Table 1.1: Anova table for one-way classification

Source	d.f.	S.S.	M.S.	F (calculated)
Factor A	s-1	$\sum_{i=1}^{s} \frac{y_{i.}^2}{n_i} - \frac{y_{..}^2}{n}$	MS_A	MS_A/MS_e
Error	n-s	By subtraction	MS_e	
Total	n-1	$\sum\sum y_{ij}^2 - \frac{y_{..}^2}{n}$		

Multiple Comparison Tests

When the treatment differences are established from the Anova table, one will be interested to isolate the treatments giving significantly different response than others. This could be achieved by multiple comparison methods and we briefly discuss some of the commonly used tests.

(i) Least significant difference (lsd): Using this test one concludes that two treatments are significantly different if the difference in their mean response exceeds the lsd value, which is the product of the tabulated t-value at the chosen level of significance based on error d.f. with the standard error of the difference of the treatment means. The lsd value for comparing the ith and jth levels of factor A which are applied to n_i and n_j experimental units is $(t_{n-s,\,\alpha/2})$. $\sqrt{MS_e\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$. The lsd value is also known as critical difference (CD). When all the levels are equally replicated, say r times, then lsd value for comparing any two different treatment means is $(t_{n-s,\,\alpha/2})$. $\sqrt{2MS_e/r}$. A bar diagram can be drawn which shows all effects which do not differ among them significantly under one bar.

Since the lsd procedure is equivalent to making many t-tests, this

method comparing different treatment means is also called multiple t-test technique.

(ii) Scheffe's Multiple Comparison Test: With this test any two treatments will be judged significant at α level of significance if the difference between their means exceeds $\{(s-1)F_{s-1,n-s,\alpha}\}^{\frac{1}{2}}$ times the standard error of the difference in their means. In other words, the ith and jth levels of factor A will have significant different effects at α level of significance if

$$|\bar{y}_i - \bar{y}_j| > \{(s-1)F_{s-1,n-s,\alpha}\}^{\frac{1}{2}} \sqrt{MS_e\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

Schelffe's multiple comparison tests are more conservative than multiple t-tests and fever treatment differences will be detected by this procedure than lsd test.

(iii) Dunnett's Multiple Comparison Test: When each of s-1 levels of factor A have to be compared with a control level of A, Dunnet proposed an alternative method rather than lsd. To perform this test, we find t' value from the table. We read the column headed by p = s-1 and row corresponding to error d.f. and choose P = .95 or .99 according as the level of significance α used is .05 or .01. Then the i^{th} level of factor A is significantly different from the control level of A if the absolute value of the difference in the mean values exceed $t'_\alpha \sqrt{2MS_e / r}$, where r is the equi-replications of all levels of factor A.

(iv) Newman–Keuls Range Test: The treatment means $\bar{y}_1, \bar{y}_2, \ldots\ldots\bar{y}_s$ will be ranked from highest to lowest as $\bar{x}_s, \bar{x}_{s-1}, \ldots, \bar{x}_1$. We now use the table. In the table, we enter s, error d.f. = n-s and protection level is the level of significance and the entry may be denoted by $q_{s,n-s,\alpha}$. Now we compute the upper α percent point of studentized critical ranges

$$W_s = (q_{s,n-s,\alpha})\sqrt{MSe / r}$$

$W_{s-1} = (q_{s-1, n-s, \alpha}) \sqrt{MSe / r}$

Compare $(\bar{x}_s - \bar{x}_1)$ with W_s

If $(\bar{x}_s - \bar{x}_1) < W_s$, we conclude that at α level of significance that all treatments are not significantly different and the process of comparison stops. However, if $(\bar{x}_s - \bar{x}_1) > W_s$, we divide $\bar{x}_s, \bar{x}_{s-1}, ..., \bar{x}_1$ into two subgroups one containing $\bar{x}_s, ..., \bar{x}_2$ and the other from $\bar{x}_{s-1}, ..., \bar{x}_1$. We compare each of $(\bar{x}_s - \bar{x}_2)$ and $(\bar{x}_{s-1} - \bar{x}_1)$ with W_{s-1}. If either range does not exceed W_{s-1}, then the means in each of the groups are equal. If either or both ranges exceeds W_{s-1}, then the s-1 means in the group concerned are divided into two groups of s-2 means each and the ranges for these subgroups are compared with W_{s-2}. This procedure is continued until a group of i means is found whose range does not exceeds W_i. In other words, by this method the difference between any two means is significant when the range of the observed means of each and every subgroup containing the two means under test is significant according to studentized critical range.

(v) Duncan's Multiple Range Test: The treatment means are ranked highest to lowest. Let these are $\bar{Y}_{s.}, \bar{Y}_{(s-1)} ..... \bar{Y}_1$, now find significant studentized ranges at 5% or 1% for multiple range test for various means and at (n-s) degree of freedom from the tables given in the books under the heading "Significant studentized ranges for 5% and 1% level Multiple Range Test'.

Let it be denoted as $(d_p, n-s, \alpha)$

Compute $D_s = (d_s, n-s, \alpha) \sqrt{MSE/r}$

$D_{s-1} = (d_{s-1}, n-s, \alpha) \sqrt{MSE/r}$, ...

We then compare $\bar{Y}_s - \bar{Y}_1$ with D_s. If $\bar{Y}_s - \bar{Y}_1 < D_s$, then we conclude that all the treatments do not differ significantly at a given level of significance otherwise differ significantly. However if $\bar{Y}_s - \bar{Y}_1 > D_s$, then we compare each of $\bar{Y}_s - \bar{Y}_2$ and $\bar{Y}_{(s-1)} - \bar{Y}_1$ with D_{s-1}. If either range is less than D_{s-1} then the treatments in that set are not

significantly different. If either range exceeds D_{s-1}, then s-1 means are divided into two sets of s-2 means each and compared with D_{s-2}. The process continues till a subset of non significant means are obtained.

As an illustration, let us consider that 4 different feeds were each fed to a lot of 5 baby chicks and the gain in weight (in kg) be as given below:

Feed	Gain in weight (in kg)	Total
A	2.75, 2.45, 2.10, 1.05, 2.60	10.95
B	3.05, 5.60, 1.50, 4.45, 3.15	17.75
C	2.10, 4.85, 4.05, 4.75, 4.60	20.35
D	8.45, 6.85, 8.45, 4.25, 7.70	35.70
		84.75

$$\text{Correction Factor} = \frac{(84.75)^2}{20} = 359.13$$

$$\text{Total sum of squares} = (2.75)^2 + (2.45)^2 + \ldots..(7.70)^2 - 359.13$$

$$= 94.48$$

$$\text{Feeds sum of squares} = \frac{10.95^2}{5} + \frac{17.75^2}{5} + \frac{20.35^2}{5} + \frac{35.70^2}{5} - 359.13 = 65.59$$

Error S.S. = 94.48-65.59 = 28.89

Anova Table

Source	d.f.	S.S.	M.S.	F
Feeds	3	65.59	21.86	12.1
Error	16	28.89	1.81	
Total	19	94.48		

The table $F_{3, 16, 0.5} = 3.24$. As the calculated F value exceeds the table value, we conclude that the feeds produce different gains in weights at 0.05 level of significance.

Multiple Comparison Tests

Least significant difference

$$\text{lsd or CD value} = (t_{n-s,\ \alpha/2}).\ \sqrt{MS_e\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}.$$

The lsd value at 0.05 level of significance for comparing the difference of means of weights for any two feeds is 2.12 $\sqrt{2 \times 1.81/5} = 1.80$.

The difference in means for feeds A and D have absolute value 4.95 which exceeds the lsd value, we conclude that feeds A and D produce significantly different gain in weights in baby chicks at 0.05 level of significance.

Scheffe's Multiple Comparison Test

Two means will differ significantly if

$$\left|\bar{y}_i - \bar{y}_j\right| > \{(s-1)F_{s-1,n-s,\alpha}\}^{\frac{1}{2}} \sqrt{MS_e\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

In our example RHS $= (3 \times 3.27)^{1/2} \sqrt{\frac{2x1.81}{5}} = 2.65$

Thus the feeds A and D significantly differ at 0.05 level of significance.

Dunnett's Multiple Comparison Test: Here two means differ significantly if $\left|\bar{y}_i - \bar{y}_j\right|$ exceed $t'_\alpha \sqrt{2MS_e / r}$ where t'α is the value from Dunnett's table for two sided comparisons.

Here $t'_\alpha \sqrt{2MS_e / r} = 2.67 \times \sqrt{\frac{2x1.81}{5}} = 2.27$

Thus the feeds A and D differ significantly at 0.05 level of significance.

Duncan's Multiple Range Test

For applying Duncan's Multiple range test, we have $\sqrt{MSe/r} = \sqrt{1.8/5} = 0.6$ and $d_{s,16,.05}$=3.23, 3.15 and 3.00 respectively for s = 4, 3 and 2. Thus $D_4 = 1.938$, $D_3 = 1.89$ and $D_2 = 1.8$.

If the treatment means are arranged in descending order, they correspond to feeds D, C, B and A. All the four means are significantly different because the difference of means of feeds D and A is 4.95 which exceed D_4. Now the three means of A, B and C are not different because the difference in means of feeds C and A is 1.88, smaller than D_3. Again the means of B, C and D are significantly different as the difference of means of feed D and B is 3.99 exceeding D_3. Thus feed D gives significant results than other feeds, while feeds A, B and C do not differ much. This state of affairs will be represented by **DCBA** showing that feeds A, B and C are at par, while feed D differ from them.

Partitioning of sum of squares through contrasts

Some times it becomes necessary to partition the S.S. due to treatment with (s-1) degree of freedom into components with single or multiple degree of freedom with meaningful interpretations. This we do with the help of contrasts.

Let T_1, T_2..... T_s are the treatment totals each based on r replications.

Then a linear function of T_1, T_2....., T_s

$Z=l_1T_1+l_2T_2 + \ldots\ldots\ldots + l_sT_s$ is called a contrast if $\Sigma l_i=0$

SS due to contrasts $Z=Z^2/\Sigma\ r_i l_i^2$ when the no. of replications for each treatment is same.

If each treatment totals are based upon r_i replication then $Z=\Sigma l_iT_i$ is a contrast if $\Sigma r_i l_i = 0$ and SS due to $Z=Z^2/\Sigma r_i l_i^2$.

Orthogonal contrasts: Two contrasts Z_1 and Z_2 where

$Z_1 = l_1T_1 + l_2T_2 \ldots\ldots\ldots\ldots\ldots + l_sT_s$, here $\sum_{i=1}^{s} l_i = 0$ $\left(\sum_{i=1}^{s} l_i = 0\right)$

$Z_2 = q_1T_1 + q_2T_2 + \ldots\ldots\ldots\ldots + q_sT_s.$ here $\sum_{i=1}^{s} q_i = 0$

are said to be orthogonal if $\Sigma\, l_i\, q_i = 0$ if treatment means are based on equal number of replications and $\Sigma\, r_i\, l_i\, q_i = 0$ if treatment means are based on unequal number of replications.

If we have s treatments the there exist (s-1) mutually orthogonal contrasts, each contrast carries one degree of freedom.

Then treatments sum of square = $\sum_{i=1}^{s-1} SS$ due to Z_i

where Z_i' s are mutually orthogonal contrasts.

In our example, we have

$T_1 = 10.95,\ T_2 = 17.75,\ T_3 = 20.35,\ T_4 = 35.70$

We have three mutually orthogonal contrasts.

$Z_1 = -3T_1 - T_2 + T_3 + 3T_4 = -3 \times 10.95 - 17.75 + 20.35 + 3 \times 35.70 = 76.85$

$Z_2 = T_1 - T_2 - T_3 + T_4 = 8.55$

$Z_3 = -T_1 + 3T_2 - 3T_3 + T_4 = 16.95$

S.S. due to $Z_1 = (76.85)^2/100 = 59.06$

S.S. due to $Z_2 = (8.55)^2/20 = 3.65$

S.S. due to $Z_3 = (16.95)^2/100 = 2.87$

Here treatment SS = 65.59 at 3 degree of freedom.

Here we see that treatments SS = $\sum_{i=1}^{3}$ SS due to Z_i

If the feeds A, B C are of one type and feed D be of another type. Then the SS due to feeds with 3 degree of freedom can be partitioned into 2 components (i) S.S. due to feeds with in the first type with 2 d.f. and (ii) S.S. between the types with 1 d.f.

S.S. due to feeds with in first type

$$= \frac{T_1^2}{r_1} + \frac{T_2^2}{r_2} + \frac{T_3^2}{r_3} - \frac{[T_1 + T_2 + T_3]^2}{r_1 + r_2 + r_3}$$

$$= [(10.95)^2 + (17.75)^2 + (20.35)^2]/5 - (49.05)^2/15 = 9.42$$

Comparison between the types of feed, we consider

$Z = T_1+T_2+T_3-3T_4 = 10.95 + 17.75 + 20.35-3 \times 35.70 = 58.05$

$$\text{S.S. due to } Z = \frac{(10.95+17.75+20.35-3\times 35.75)^2}{5(1+1+1+9)} = 56.16$$

The total of these two sums of squares adds upto the feed sum of squares.

Example

Consider the following data

Treatment (varieties)

	1	2	3	4
	45	35	34	41
	46	33	34	41
	49		35	44
	44		34	43
			33	41
				42
				44
				41
				41
Total	184	68	170	378
Means	46	34	34	42

Here $r_1 = 4$, $r_2 = 2$, $r_3 = 5$, $r_4 = 9$, GT = 800

Suppose varieties 1 and 2 are from one origin and 3 and 4 from other origin. It would be desirable to have comparisons of

(i) Varieties 1 and 2 with 3 and 4

(ii) Variety 1 with 2

(iii) Variety 3 and 4

Solution

Then desired contrasts will be

$$Z_1 = 7(T_1 + T_2) + (-3)(T_3 + T_4)$$

$$= 7T_1 + 7T_2 - 3T_3 - 3T_4$$

$$Z_2 = 1T_1 + (-2)T_2 + 0T_3 + 0T_4 = T_1 - 2T_2$$

$$Z3 = 0T_1 + 0T_2 + 9T_3 + (-5)T_4$$

$$= 9T_3 - 5T_4$$

Here the treatment totals depends upon unequal no. of observations. In the first contrast we want to compare the mean of 4 + 2 = 6 observations with mean of 5 + 9 = 14 observations. So it is necessary to have some weighing for comparison of treatment totals based on unequal number of observations. Since the smallest integer may be divided evenly by both 6 and 14 is 42, we see that 7 and 3 are the indicated weights to be used. In the same way coefficient of Z_2 and Z_3 are found out and it should be verified that all those three contrast should be mutually orthogonal to each other. Then only the total treatment sum of square will be equal to the total of the sum of square due to contrasts.

The S.S. for these contrast can be obtained as under.

$$S.S.(Z_1) = \frac{[7(184) + 7(68) + (-3)(170) + (-3)(378)]^2}{[4(7^2) + 2(7^2) + 5(-3)^2 + 9(-3)^2]} = 34.3$$

$$S.S.(Z_2) = \frac{[(1)(184) + (-2)(68) + (0)(170) + 0(378)]^2}{[4(1^2) + 2(-2)^2 + 5(0)^2 + 9(0)^2]} = 192.0$$

$$S.S.(Z_3) = \frac{[(0)(184) + (0(68) + 9(170) + (-5)(378)]^2}{[4(0^2) + 2(0)^2 + 5(9)^2 + 9(-5)^2]} = 205.7$$

The above S.S. can also be obtained as under

$SS(Z_1) = (184 + 68)^2/16 + (170 + 378)^2/14 - (800)^2/20 = 34.3$

$SS(Z_2) = (184)^2/4 + (68)^2/2 - (252)^2/16 = 192.0$

$SS(Z3) = (170)^2/5 + (378)^2/9 - (548)^2/14 = 205.7$

Now ANOVA table can be made as under:

ANOVA Table

Source of variation	d.f.	SS	MS	F
Treatments	3	432	144	72
1 and 2 vs 3 and 4	1	34.3	34.7	17.15
1 Vs 2	1	192.0	192.0	96.0
3 Vs 4	1	205.7	205.7	102.85
Error	16	32	2.0	

We find all above difference are significant

2. Random Effects Model (Variance Component Model):

One way classification: Let the s levels of a factor be chosen randomly from all possible levels of that factor and let a_i be the effect of the ith chosen level. Since the experimented levels are randomly selected, the experimenter will be mainly interested to test the equality of the effects for different levels in the population rather than the equality of effects of the tested levels. Since the equality of a's implies that the variance of a's in the population σ_A^2 is zero, with a random effects model the null hypothesis to be tested will be $H_0 = \sigma_A^2 = 0$.

Let the s selected levels of factor A be experimented on n units, he ith level being randomly applied to n_i units for i = 1, 2, ..., s $\left(\sum_{i=1}^{s} n_i = n\right)$. The model assumed will be.

$$y_{ij} = \mu + a_i + e_{ij}, \quad j = 1, 2, \ldots\ldots n_i; \quad i = 1, 2, \ldots\ldots, s$$

where y_{ij} is the observation on the jth unit to which the ith level of A is applied, μ is the general mean, a_i is the effect of the ith level of factor A and e_{ij}'s are random errors. It is assumed that a_i's are normally and independently distributed with mean zero and variance σ_A^2 ; e_{ij}'s are normally and independently distributed with mean zero and variance σ_e^2 ; a_i's are independently distributed with e_{ij}'s. The ANOVA will be formed as in the case of fixed effect model. The average value of mean square for all sampling variations in a's and 's will be called the expected value of mean squares and will be

denoted by E(M.S.) for a one way classification as under:

Anova Table for a Random Effects model for one way classification.

Source	d.f.	S.S.	M.S.	E(M.S.)
Factor A	s-1	$\sum \frac{Y_i^2}{ni} - \frac{Y_{..}^2}{n}$	MS_A	$\sigma_e^2 + \frac{1}{s-1}\left(n - \frac{\sum n_i^2}{n}\right)\sigma_A^2$
Error	n-s	By subtraction	MS_e	σ_e^2
Total	n-1	$\sum_{i,j} y_{ij}^2 - \frac{y_{..}^2}{n}$		

$$\hat{\sigma}_e^2 = MS_e,\ \hat{\sigma}_A^2 = \frac{MS_A - MS_e}{\frac{1}{s-1}\left(n - \frac{\sum n_i^2}{n}\right)}$$

If n_i's are all equal to r, say, then $\hat{\sigma}_A^2 = \frac{MS_A - MS_e}{r}$ under the null hypothesis $H_0 : \sigma_A^2 = 0$, MS_A and MS_e will both independently estimate σ_e^2 and this leads to the following:

If $MS_A/MS_e > F_{s-1,\ n-s,\ \alpha}$, we reject the null hypothesis H_0: $\sigma_A^2 = 0$ and conclude that H_1: $\sigma_A^2 \neq 0$ at α level of significance; otherwise not.

If the feeds of the data of earlier example were randomly taken from a populations of feeds with σ_A^2 as the variance of the effects of the feeds, then $\hat{\sigma}_A^2$ will be unbiasedly estimated by

$$\hat{\sigma}_A^2 = \frac{21.86 - 1.81}{5} = 4.01$$

One unhappy situation that arises in the variance component estimation is that one may get negative estimates for variances and in such cases, one may conclude that there is stronger evidence for the corresponding population variances to be closer to zero.

❐ ❐ ❐

Chapter – 3

Transformation of Data

The interpretation of data based on analysis of variance (ANOVA) is valid only when the following assumptions are satisfied.

1. Additive effects: Treatment effects and the block effect are additive.
2. Independence of Errors: Experimental errors are independent.
3. Homogeneity of variance: Observations have common variance.
4. Normal Distribution: Character under study follows normal distribution.

Also the statistical tests t, F, Z are valid under the assumption of independence of errors and normality of characters under study.

There are certain types of data where all these assumptions are not satisfied. For example, when the data are in the form of percentage say, seeds that germinate in a plot, the number of rare insects present in soil samples or the number of particular types of impurities in milk sample, the assumption that the variance is constant for the different observations, does not hold good. In the first case, the observations have Binomial distribution and hence the variance which depends upon the unknown percentage is not a constant. In other cases, the observations have a Poisson distribution. In such cases, a non-linear transformation of the observations satisfies all the above assumptions. For example

Model	Additive		Multiplicative		$Log_{10}X$	
Block	1	2	1	2	1	2
Treatment -1	10	20	10	20	1	1.30
Treatment - 2	30	40	30	60	1.48	1.78
Treatment effect	20	20	20	40		

It is seen with the help of the log transformation, multiplicative model can be changed into additive model. Another hypothetical set of data with multiplicative effects of treatment and replication is

Treatment	Replication		Replication effect	
	I	II	I-II	100 (1-II) / II
A	200 (2.30103)	125 (2.09691)	75 (0.20412)	60
B	160 (2.20412)	100 (2.0000)	60 (0.20412)	60
Treatment effect (A-B)	40 (0.09691)	25 (0.09691)		
100(A-B)/B	25	25		

In this case the treatment effect is not constant over replication and the replication effect is also not constant over treatments. However, when both treatment effect and replication effect are present in percentages, their effects become additive with the logarithmic transformation y = logx and the replication and treatment effect becomes additive. Logarithmic transformed values are given in bracket.

Types of Transformation

Square root transformation

Firstly, mean and variance for each treatment may be worked out and graph between means and variances may be drawn. In case we

get a straight line we conclude that variance is varying with mean. Here the original data follows Poisson distribution where the mean and variance are equal, so the square root transformation is useful in bringing the original distribution into normal distribution.

For example, the number of plants infected with a particular disease in a given area, the number of insects of a particular disease in a given area, the number of weed seeds in a sample of seed, the bacterial colonies on a plate count etc. follow Poisson distribution. In this case $\sqrt{y}$ transformation is recommended for y, when the original values lies between 0 to 30 or 70 to 100. However when the original data consists of zeros then $\sqrt{\left(y+\frac{1}{2}\right)}$ or $\sqrt{(y+1)}$ or $\sqrt{\left(y+\frac{3}{8}\right)}$ transformation may be used. Here the square root transformation approximates Poisson to Normal distribution.

Angular Transformation (Sine Inverse Transformation, or Arc Sine Transformation)

When data relate to Binomial distribution, mean is proportional to variance, angular transformation (also termed as sine inverse transformation) as given by the equation. $\theta = \sin^{-1} X$ may be used to stabilize the variance and transformed data follows Normal distribution. Here θ is in degrees and X is the percentage value. Angles corresponding to percentage, Angle = ARC SINE $\sqrt{\text{Percentage}}$ as given by C.I. BLISS are given in Fisher and Yates statistical tables. When all the percentages lies between 30 to 70 the data are expected to follow normal distribution and hence no transformation is required. The zero percent should be substituted by (1/4n) and 100 by $\left(100-\frac{1}{4n}\right)$ where n is the no. of units upon which percentage of data was based.

Here not all percentage data need to be transformed and even if they do Arc Sine transformation is not the only transformation possible. The following rules may be useful in choosing the proper transformation, scale for percentage data derived from count data.

Rule 1. The percentage data lying between the range of 30 to 70% is homogeneous and no transformation is needed

Rule 2. For percentage data lying within the range of either 0 to 30% or 70 to 100% but not both, the square root transformation should be used.

Rule 3. For percentage that do not follow the ranges specified in Rule 1 or Rule 2, the Arc Sine transformation should be used.

Logarithmic Transformation

When mean is proportional to standard deviation, or in other words, if the coefficient of variation is constant, the transformation y= log x is used to stabilize the variance. If zero occurs or data sheet involves small values (e.g. less than 10) log (y+1) or log $\left(y + \frac{3}{8}\right)$ is used. This transformation is particularly useful when the observed data consists of big integers such as Index numbers or biological population or the population following multiplicative model for example numbers of insects per plant, number of egg masses per plant or number of leaving larvae on rice plants treated with various rates of an insecticides are typical examples.

This transformation would make errors normal when observations follow negative binomial distribution.

When the data relates to proportions we can use the transformation y=log x/(1-x) when x represents the proportion. For example, A_1 denotes the area of a plot affected by some disease and A is the total area of the plot then $x = A_1/A$ (proportion of the area affected) and will lie between 0 and 1.

Reciprocal transformation

When variances differs from treatment to treatment, if time is characteristic under study, reciprocal transformation converting the original values to reciprocals, would bring stabilization to the variances of the treatments.

Probit Transformation

This transformation is often used in the analysis of dose-response relationship when the response variable is a binomial proportion $\frac{r}{n}$. It is useful when, for example, the threshold of response is normally distributed in the test population because in this case it reduces the dose response relationship to a straight line. The probit corresponding to r/n is defined as (X+5) where X is a standardized normal deviate such that $\phi(X)=\frac{r}{n}$. The constant 5 is added so that the probit is always positive. Thus of r=7, n = 8, then $\frac{r}{n}$ = 0.875, from table it is seen that for x = 1.15, $\phi(X)$ = 0.875. Hence the probit corresponding $\frac{r}{n}$ = 6.75. Table giving the probit transformation directly are available.

ANOVA table is constructed as usual manner as the transformed data. For overall comparison of means, F test is done as usual with the transformed data. Similarly, for pairwise comparison of means, LSD or DMRT are done with the transformed data. Of course, the final presentation of the treatment means in the original scale, they may either be computed from the original values or they may be obtained by converting the transformed values to the original scale.

In certain cases it may so happen that the ranking of means scale and the transformed scale are not same. In such situations, computation of means by converting the transformed data is more appropriate.

Transformation affects the apparent importance of interactions. In some cases, while un-transformed data show no existence of interactions, the transformed data may show existence of interaction. In such cases, the interpretation becomes difficult and analysis of data with the original scale is rather better.

Example

The data on percentage of insect survival in 12 rice varieties trial

in a completely randomised design with three replications is given below. For each plant out of so cased insets were different of surviving percentage. Analyze the data using appropriate transformation.

Percentage Survival

Variety	Rep-I	Rep-II	Rep-III	Total	Mean.
V_1	37.00	25.33	47.00	109.33	36.44
V_2	21.33	47.32	81.00	149.65	49.88
V_3	0.00	0.00	0.00	0.00	0.00
V_4	25.32	27.88	51.10	104.30	34.77
V_5	23.89	27.32	52.33	103.54	34.51
V_6	0.00	0.00	21.00	21.01	7.00
V_7	32.82	28.67	25.89	87.38	29.13
V_8	0.00	0.00	0.00	0.00	0.00
V_9	18.89	35.48	12.75	66.79	22.26
V_{10}	95.89	100.00	100.00	295.89	98.63
V_{11}	13.56	32.34	38.58	84.48	28.16
V_{12}	44.33	48.66	12.15	105.14	35.05

Solution:

The arc sine transformation should be used because the percentage data range from 0 to 100%.Before transformation all zero per cent values are replaced by $\frac{1}{4n}$ i.e. $\left[\frac{1}{4x75}\right]$ and all 100 per cent values by $\left(100-\frac{1}{4n}\right)$ i.e. [100-1/4(75)]. The transformed data and the analysis of variance is given below. LSD test and the DMRT were first applied to the transformed means and then transformed to the original means.

Transformation of data using Arc sine Transformation:

Variety	Transformed values			Total	Mean
	Rep-I	Rep-II	Rep-III		
V_1	37.46	30.22	43.28	110.96	36.99
V_2	27.51	43.46	64.16	135.13	45.04
V_3	0.33	0.33	0.33	0.99	0.33
V_4	30.21	31.87	45.63	107.71	35.90
V_5	29.26	31.51	46.34	107.11	35.70
V_6	0.33	0.33	27.27	27.94	9.31
V_7	34.95	32.37	30.59	97.91	32.64
V_8	0.33	0.33	0.33	0.99	0.33
V_9	25.52	36.56	20.92	83.00	27.67
V_{10}	78.30	89.67	89.67	257.64	85.88
V_{11}	24.61	34.66	38.40	94.66	31.55
V_{12}	41.74	44.23	20.40	106.38	35.46

Analysis of variance table

Source of variation	Degree of freedom	Sum of squares	Mean squares	Computed F
Variety	11	17024.13	1547.65	16.26**
Error	24	2284.09	95.17	
Total	35	19308.22	551.66	

** Significant at 1% level.

$$LSD_{.05} = 2.03 \times \sqrt{\frac{2x95.17}{3}} = 18.32$$

$$LSD_{.01} = 2.73 \times \sqrt{\frac{2x95.17}{3}} = 21.74$$

These LSD values are in the transformed scale and must be applied to the transformed means. For example to compare first and second variety, the difference of the transformed means is 45.04 – 36.99 = 8.05 which is less than the LSD value 18.32 at 5% level of significance.Hence variety 1 and variety 2 do not differ significantly.

❐ ❐ ❐

Chapter – 4

Single Factor Experiments

Experiments in which only a single factor varies while all others are kept constant are called single factor experiments. For example (i) Fertilizer trials where several rates of single fertilizer element are tested. (ii) Insecticides trials where several insecticides are tested. (iii) Plant population trials where several plant densities are tested.

There are two groups of experimental situations that are applicable to a single factor experiment. One group is the family of complete block designs which is suited for treatments with a small number of treatments and is characterized by blocks, each of which contains at least one complete set of treatments. The other group is the family of incomplete blocks which is suited for experiments with a large number of treatments and is characterized by blocks, each of which contains only a fraction of the treatments to be tested.

Hence three complete block designs namely Completely Randomized, Randomized Complete Block and Latin Square Design will be discussed. The other two incomplete block designs namely Balanced Incomplete Block Design and Lattice Design will be discussed in the later chapters.

Completely Randomized Design (CRD)

Designs are usually characterized by nature of grouping of experimental units and the procedure of random allocation of

treatment to the experimental units. In a completely randomized design the units are taken in a single group and as far as possible the unit forming the group should be homogenous in nature. For example all the field plot constituting the group are having the same soil fertility, soil depth, soil texture, soil moisture, etc., all the cows forming the group are of the same breed, same age, same weight, same lactation etc. Suppose there are v treatments and the i^{th} treatment is replicated r_i times thus we require $\sum_{i=1}^{v} r_i = n$ experimental units. Allote v treatment randomly to n experimental units such that the i^{th} treatment occupies r_i units or plots. Such a design is called CRD. This design is also seldom called a non-restrictional design. CRD is most useful for laboratory experiments, pot experiments, green house experiments, and sometimes for poultry and animal experiments.

Layout

Suppose there are four treatments T_1, T_2, T_3 and T_4 which are replicated 4, 3, 3 and 5 times respectively. For layout in CRD the first treatment will be allotted at random to 4 units, second and third treatment to 3 units and fourth treatment to 5 units of 15 experimental units.

Field plan of CRD with treatments assigned to plots.

T_2 (1)	T_2 (6)	T_2 (11)
T_3 (2)	T_4 (7)	T_3 (12)
T_4 (3)	T_4 (8)	T_1 (13)
T_1 (4)	T_1 (9)	T_1 (14)
T_3 (5)	T_4 (10)	T_4 (15)

Merits of CRD

(i) Its layout is very simple.

(ii) There is complete flexibility in this design i.e. any number of treatments and replicates for each treatment can be taken.

(iii) Whole experimental material can be utilized in this design.

(iv) The analysis is not complicated by missing data.

(v) The statistical analysis is not complicated by unequal replication of treatments. This is not the case with more complex designs.

(vi) The design offers maximum error degree of freedom. No other design with the same total number of plots and treatments provides more error degree of freedom than does the CRD.

Demerits of CRD

CRD suffers from many lacunae which are as follows

(i) The design is suitable for a small number of treatments.

(ii) It is difficult to find homogenous experimental units in all respect.

(iii) It is seldom suitable for field experiments as compared to other experimental designs.

Analysis

This design provides a one way classified data according to the levels of a single factor. For its analysis following model is used

$Y_{ij} = \mu + t_i + e_{ij}$ $\quad i = 1, 2, \ldots\ldots\ldots v;$ $\quad j = 1, 2, \ldots\ldots r_i$

Set up a table of observations and compute the treatment totals.

Treatments					**Total**	**Mean**
1	Y_{11}	Y_{21}	.....	Y_1r_1	T_1	$\bar{y}_1$
2	Y_{12}	Y_{22}	.....	Y_2r_2	T_2	$\bar{y}_2$
.	.					
.	.					
v	Y_{v1}	Y_{v2}		Y_vr_v	T_v	$\bar{y}_v$
Total					G	$\bar{\bar{y}}$

where Y_{ij} is the observation obtained from the j^{th} replicate of the

i^{th} treatment, μ is the general mean, t_i is the fixed effect of the i^{th} treatment and e_{ij} is the error component which is random variable assumed to be normally and independently distributed with zero mean and a constant variance σ^2.

Here $\sum_{j=1}^{r_i} Y_{ij} = T_i$ = Sum of the observations of the i^{th} treatment

$\bar{y}_i = T_i/r_i$ = mean of the i^{th} treatment

$G = \sum_{i=1}^{v} T_i$ = grand total of all the observation

$\bar{\bar{y}} = G/n$ = grand mean; here n is the total no. of observation.

Compute correction factor

$CF = G^2/n$

Then total sum of squares (SSTOT) $= \sum_i \sum_j Y_{ij}^2 - CF$

Treatment sum of squares (SST) $= \sum_{i=1}^{v} T_i^2/r_i - CF$

Error sum of squares (SSE) = Total SS – Treatment SS

ANOVA

Source	d.f.	SS	M.S.S.	F
Treatments	v -1	SST	SST/v-1 =MST	MST/MSE~F[v-1, (n-v)]
Error	(n - v)	SSE	SSE/(n-v) = MSE	
Total	n-1		SSOT	

Where MST and MSE are the mean squares for treatments and error respectively. If the calculated vale of F is greater than the table value [α, (v-1), n-v] where α denotes the level of significance. The Hypothesis H_o: $\mu_1 = \mu_2 \ldots\ldots = \mu_v$ (the true mean of all treatments are the same) against H_0 : $\mu_1 \neq \mu_2 \ldots\ldots \neq \mu_v$ (at least one mean differs

from the others) is rejected and it can be inferred that any two treatment effects are significantly different from each other.

SE of any treatment mean say i^{th} treatment mean is

$$SE_m = \sqrt{\frac{MSE}{r_i}}$$

SE of difference between any two treatment means is

$$SE_d = \sqrt{MSE\left(\frac{1}{r_i}+\frac{1}{r_j}\right)}$$ where r_i and r_j are the number of replications for the i^{th} and j^{th} treatment.

If $r_i = r_j = r$ then

$$SE_d = \sqrt{\frac{2MSE}{r}}$$

The critical difference (C.D.) can be calculated as

$$C.D. = SE_d \times t_{(\alpha,\ n-v)}$$

Where $t_{(\alpha,\ n-v)}$ is the two tailed value of 't' at α level of significance and at (n-v) error degree of freedom.

Means are arranged in ascending or descending order of magnitude for comparison. Any two treatments means are said to differ significantly if there difference is larger than the critical difference.

Example

The yield of Lentill (Kg/plot) from 20 demonstration plots on various soil types is given below. Use CRD for analysis to see if there were yield differences due to soil type. Lentill seed yields (Kg/plot) from 20 demonstration plots on various soil types are given in following table.

Analysis

$$CF = G^2/n = (546.9)^2/20 = 14954.98$$

Total sum of squares = $\sum_i \sum_j Y_{ij}^2 - CF$

= $(41.3)^2 + (33.7)^2 \ldots\ldots\ldots\ldots + (29.5)^2 - CF$

= $16217.25 - 14954.98 = 1262.27$

Soil type

Treat-ments	1	2	3	4	5	Total
	41.3	29.2	19.7	41	31.5	
	33.7	27.5	20.6	36.9	25.7	
	28.9	23.4	14.1	32.0	29.5	
		20.3	11	32.8		
		19.9		27.9		
Sum	103.9	120.3	65.4	170.6	86.7	546.9
r_i	3	5	4	5	3	20
Mean	34.63	24.06	16.35	34.12	28.90	27.34

Treatment sum of square = $\frac{\Sigma T_i^2}{r_i} - CF$

= $\frac{(103.9)^2}{3} + \ldots\ldots + \frac{(86.7)^2}{3} - CF$

= 15888.61-14954.98 = 933.63

Sum of square due to Error

= Total SS – Treat SS

= 1262.27 -933.63 =328.64

ANOVA Table

Source	d.f.	SS	MS	F
Soil type	4	933.63	233.41	10.65
Error	15	328.64	21.91	
Total	19	1269.27		

** Significant at the 1% level

The Table value of $F_{.01}(4,15) = 4.8932 < 10.65$

which means the treatments differ significantly

Standard Error of mean = $\sqrt{\frac{MSE}{r_i}}$

$$SE_{\bar{Y}_1} = \sqrt{\frac{21.91}{3}} = 2.70$$

$$SE_{\bar{Y}_2} = \sqrt{\frac{21.91}{5}} = 2.09$$

$$SE_{\bar{Y}_3} = \sqrt{\frac{21.91}{4}} = 2.34$$

$$SE_{\bar{Y}_4} = \sqrt{\frac{21.91}{5}} = 2.09$$

$$SE_{\bar{Y}_5} = \sqrt{\frac{21.91}{3}} = 2.70$$

SE of difference between type 1 and type 2 soil.

$$SE_{(\bar{y}_1-\bar{y}_2)} = \sqrt{MSE\left(\frac{1}{r_i}+\frac{1}{r_j}\right)} = \sqrt{21.91\left(\frac{1}{3}+\frac{1}{5}\right)} = 3.42$$

$$SE_{(\bar{y}_1-\bar{y}_3)} = \sqrt{21.91\left(\frac{1}{3}+\frac{1}{4}\right)} = 3.57$$

$$SE_{(\bar{y}_1-\bar{y}_4)} = \sqrt{21.91\left(\frac{1}{3}+\frac{1}{5}\right)} = 3.42$$

$$SE_{(\bar{y}_1-\bar{y}_5)} = \sqrt{21.91\left(\frac{1}{3}+\frac{1}{3}\right)} = 3.82$$

$$SE_{(\bar{y}_2-\bar{y}_3)} = \sqrt{21.91\left(\frac{1}{5}+\frac{1}{4}\right)} = 3.14$$

$$SE_{(\bar{y}_2-\bar{y}_4)} = \sqrt{21.91\left(\frac{1}{5}+\frac{1}{5}\right)} = 2.96$$

$$SE_{(\bar{y}_2-\bar{y}_5)} = \sqrt{21.91\left(\frac{1}{5}+\frac{1}{5}\right)} = 3.42$$

$$SE_{(\bar{y}_3-\bar{y}_4)} = \sqrt{21.91\left(\frac{1}{4}+\frac{1}{5}\right)} = 3.14$$

$$SE_{(\bar{y}_3-\bar{y}_5)} = \sqrt{21.91\left(\frac{1}{4}+\frac{1}{5}\right)} = 3.58$$

$$SE_{(\bar{y}_4-\bar{y}_5)} = \sqrt{21.91\left(\frac{1}{5}+\frac{1}{5}\right)} = 3.42$$

C.D. values for comparing the different pairs of treatment can be calculated. For example for comparing $\bar{y}_1$, with $\bar{y}_2$ we had

CD at 1% level of significance = $SE_{(\bar{Y}_1-\bar{Y}_1)} \times t_{(.01,15)}$

$= 3.42 \times 2.947 = 10.08$

$|\bar{y}_1 - \bar{y}_2| = |34.63 - 24.06| = 10.57 > 1.55$ So the means $\bar{y}_1$ and $\bar{y}_2$ differ significantly. In the same way CD for all the pairs can be calculated at the desired level of significance.

Analysis of yield data from 20 demonstration plots indicates that there are highly significant differences in yield among the five soil types represented in the trial. The results are presented in the following table.

Mean yield (kg/plot) of wheat grown on five soil types

Soil types	1	2	3	4	5
Mean yield	34.63	24.06	16.35	34.12	28.90
Replication	3	5	4	5	3
Standard Error	2.70	2.09	2.34	2.09	2.70

It is apparent that soil type-I produce the highest yield of lentil seed. Soil type 3 is clearly inferior to others. The soil types can be compared with the CD values.

Randomized Block Design

The Randomized block design (RBD) or more, correctly the randomized complete block design is the design used most often in agricultural research. When the experimental material is not entirely homogeneous, we divide the material in to homogeneous groups of the experimental units by adopting the principle of local control. The object is to make the variation from plot to plot as small as possible with in the blocks while maximizing the variation among blocks. Changes in technique or in other condition that might affect the results should be made between blocks. The number of blocks in the design is the same as the number of replications.

Suppose we want to compare the effect of v treatment, each treatment being replicated an equal number of times say 'r'. We need n=rt experimental units. The first step is to divide the units in to r homogeneous blocks each consisting of v experimental units. Then the second step is to allot the treatments randomly to the units of the block. The randomization is to be done afresh for each block.

For actual layout in the field, blocks are formed perpendicular to the direction of fertility gradient in the field and then blots with in each block are formed in the direction of fertility gradient. All block and plots must have the same size.

Layout of RBD

Let there be five treatments A, B, C, D and E each replicated 5 times. The experiment area (fertility of the area having a slope from x to y) can be divided in to five blocks perpendicular to the direction of fertility gradient as shown below and then each block is divided in to five plots for the random allotment of the treatments.

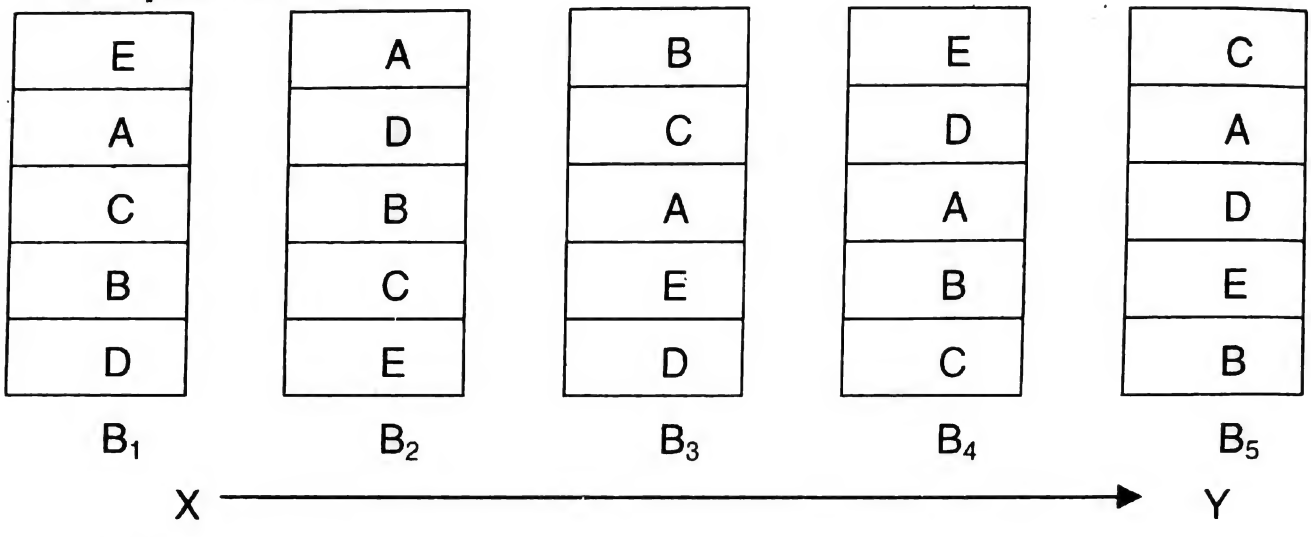
Salient features of RBD

1. It can remove one source of variation from the experimental error and thus increased precision.
2. By blocking blocks under different conditions it can broaden

the scope of the trial.

3. Any number of treatments and any number of blocks may be used. The only restriction is that each treatment must be replicated the same number of times in each block.

Field plan of RBD



4. Statistical analysis is fairly simple.

5. Even some treatments are missing in one or more blocks, the statistical analysis can be performed with the help of missing plot technique. If the numbers of missing units are more in any particular block, the statistical analysis can still be done by removing the entire block from the analysis.

6. In clinical or similar trials where animals like rats, guinea pigs etc., are the experimental units, animals coming from the same litter may form blocks. Characters like age, sex, vigour, breed etc., could be considered for forming blocks. Milk of the same fat contents would be considered as blocks in dairy experiments. Shelves of the germinators could be blocks in horticulture experiments, soils of the same type would from block in soil science experiments.

   In the green house experiments, if the access of air current and sunlight are across the bench then the block would be formed along the bench and vice – versa

7. The design is less efficient than others in the presence of more than one source of unwanted variation.

Analysis

The data collected from experiments with randomized block design form a two way classification that is, classified according to the levels of two factors, viz. blocks and treatment. There are vr cells in the two way table with one observation per cell.

The model used is

$Y_{ij} = \mu + t_i + b_j + e_{ij}$ $\quad i = 1, 2, \ldots. v; \quad j = 1, 2 \ldots. r$

where Y_{ij} denotes the observation for the i^{th} treatment in the i^{th} block. μ, t_i, b_j are respectively the general mean, effect of i^{th} treatment and effect of j^{th} block. These effects are fixed and e_{ij} is the random error component which is distributly normally and independently with zero mean and constant variance σ^2.

The data can be arranged in the following two way classification table.

Blocks

	1	2		j		r	Total
Treatment 1	Y_{11}	Y_{12}		Y_{1j}		Y_{1r}	$Y_{1.} = T_1$
2	Y_{21}	Y_{22}		Y_{2j}		Y_{2r}	$Y_{2.} = T_2$
.							
.							
i	Y_{i1}	Y_{i2}		Y_{ij}		Y_{ir}	$Y_{i.} = T_i$
.							
.							
v	Y_{v1}	Y_{v2}				Y_{vr}	$Y_{v.} = T_v$
Total	$Y.1 = B_1$	$Y.2 = B_2$		$Y.j = B_j$		$Y.r = B_r$	G

Let $\sum_j Y_{ij} = Y_{i.} = T_i (i = 1,2...v)$ = Total of observation from the i^{th} treatment

and $\sum_i Y_{ij} = Y_{.j} = B_j (j = 1,2...r)$ = Total of observations from j^{th} block.

Further $\Sigma T_i = \Sigma B_j = G$

Correction factor (CF) = $\frac{G^2}{rv}$.

Sum of squares due to Treatments = $\frac{\Sigma T_i^2}{r} - CF$

Sum of square due to blocks = $\sum_j \frac{B_j^2}{v} - CF$

Total sum of equal = $\sum_{i,j} Y_{ij}^2 - CF$

Sum of square due to Error = Total SS-SS due to treatment - SS due to blocks

Here Null hypothesis H_0: $t_1 = t_2 \ldots\ldots = t_v$ against the alternative that any two t's are not equal

Analysis of variance table (ANOVA)

Source	d.f.	S.S.	MS	F
Blocks	r-1	SSB = $\sum_j \frac{B_j^2}{v} - CF$	MSB = SSB/(r-1)	$\frac{MSB}{MSE} \sim F_{[(r-1),\,(r-1)\,(v-1)]}$
Treatment	v-1	SST = $\frac{\Sigma T_i^2}{r} - CF$	MST = SST/(v-1)	$\frac{MST}{MSE} \sim F_{[(v-1),\,(r-1)\,(v-1)]}$
Error	(r-1) (v-1)	SSE= By subtraction	MSE=SSE/(v-1)(r-1)	
Total	vr-1	$\sum_i \sum_j y_{ij}^2 - CF$		

If $F_{cal} > F[\alpha; (v-1), (v-1)\,(r-1)]$ then H_0 is rejected for treatment effects.

Standard Errors

1. The Standard Error of treatment means $SE_m = \sqrt{\frac{MSE}{r}}$

2. The standard error of difference between any two treatment means

$$SE_d = \sqrt{\frac{2MSE}{r}}$$

Critical difference at α level of significance

$CD = SE_d \times [t_\alpha, (r-1)(v-1)]$.

If the difference of any two treatment means is greater than the CD value, the corresponding treatment effects are significantly different.

Example

An agronomist wanted to determine the effect of different sources of nitrogen on the yield of barely used as a forage crop. There are five sources of nitrogen.

(1) $(NH_4)_2 SO_4$ (2) $NH_4 HO_3$ (3) $CO(NH_2)_2$ (4) $Ca(NO_3)_2$ and (5) $NANO_3$. He also decided to use (6) control treatment with no nitrogen. The field plan and yields (kg/plot) are given below. Analyze the data using RBD

Soil type

I	II	III	IV
(5) 27.2	(1) 36.5	(6) 27.2	(4) 32.3
(3) 26.3	(2) 30.7	(1) 40.7	(2) 31.8
(6) 23.2	(6) 24.1	(5) 32.8	(6) 25.9
(2) 30.6	(5) 32.0	(4) 36.5	(1) 36.3
(4) 25.4	(3) 28.3	(3) 33.1	(5) 32.2
(1) 31.5	(4) 32.1	(2) 38.5	(3) 32.1

Analysis

Prepare two way table of yield (kg/plot) of Forage barley under various sources of nitrogen.

Treatment	Soil type				Total	Mean
	I	II	III	IV		
1	31.5	36.5	40.7	36.3	145.0	36.25
2	30.6	30.7	38.5	31.8	131.6	32.90
3	26.3	28.3	33.1	32.1	119.8	29.95
4	25.4	32.1	36.5	32.3	126.3	31.57
5	27.2	31.0	32.8	32.2	124.2	31.05
6	23.2	24.8	27.2	25.9	100.4	25.10
Total	164.2	183.3	208.8	190.6	747.3	
Mean	27.37	30.53	34.80	31.77		31.14

Computations

$$CF = \frac{G^2}{rv} = \frac{(747.3)^2}{24} = 23269.05$$

Sum of squares due to treatment = $\frac{\Sigma T_i^2}{r} - CF$

$$= \frac{[(145.0)^2 + .. + (100.4)^2]}{4} - CF$$

$$= 23538.27\text{-}23269.05 = 269.22$$

Sum of squares due to blocks = $\Sigma B_j^2/v - CF$

$$= \frac{[(164.2)^2 + .. + (190.6)^2]}{6} - CF$$

$$= 23438.85 - 23269.05 = 169.80$$

Total sum of square = $\Sigma\Sigma Y_{ij}^2 - CF$

$$= (31.5)^2 + (30.62)^2 + ... + (25.9)^2 - CF$$

$$= 23746.95 - 23269.05 = 477.90$$

Sum of squares due to Error = Total SS – SS due to treatment – SS due to blocks

$$= 477.90 - 269.22 - 169.80 = 38.88$$

ANOVA Table

Source	d.f.	SS	MS	F
Soil (Blocks)	3	169.80	56.60	21.84**
Fertilizer (Treatment)	5	269.22	53.84	20.78**
Error	15	38.88	2.59	
Total	23	477.90		

** Significant at 1 percent level.

Here tabulated $F_{0.01(5,15)} = 4.56$ Since $20.78 > 4.56$

We conclude that differences among the nitrogen sources means are significant at 1% level.

Again tabulated $F_{0.01(3,15)} = 5.42$ since $21.84 > 5.42$

We conclude that blocking on soil type was effective in reducing the experimental error.

1. Standard Errors

$$SE_m = \sqrt{\frac{MSE}{r}} = \sqrt{\frac{2.59}{4}} = 0.80$$

2. Standard error of a difference between two treatment mean

$$SE_d = \sqrt{\frac{2MSE}{r}} = \sqrt{\frac{2 \times 2.59}{4}} = 1.14$$

CD at 1% = SE_d x t (.01, 15)

= 1.14 x 2.95 = 3.36

Mean yield (kg/plot) of different treatments.

Source	1	2	3	4	5	6	SE_m	CD 1%
Mean	36.3	31.8	32.1	32.3	32.2.	25.9	0.80	3.36

Relative efficiency of RBD vs CRD

$$\text{Efficiency} = \frac{\text{Amount of information in RBD}}{\text{Amount of information in CRD}}$$

$$\text{Amount of information for any design} = \frac{(n_e + 1)}{(n_e + 3)} \times \frac{1}{\text{Error mean square}}$$

where n_e is the degree of freedom in the design for error.

If RBD were conducted in CRD i.e. if the blocks were not formed than

$$\text{Error mean square in CRD} = \frac{n_b E_b + (n_t + n_e)E_e}{n_b + n_t + n_e}$$

$$= \frac{3 \times 56.60 + (5+15) \times 2.59}{3+5+15} = 9.64$$

where n_b, n_t and n_e are the degree of freedom of blocks, treatments and error respectively and E_b and E_e are the block and error Mean squares

$$\text{Amount of information in CRD} = \frac{(n_e+1)}{(n_e+3)} \times \frac{1}{\text{Error mean square}}$$

$$= \frac{(18+1)}{(18+3)} \times \frac{1}{9.64} = 0.0939$$

$$\text{Amount information in RBD} = \frac{(n_e+1)}{(n_e+3)} \times \frac{1}{\text{Error mean square}}$$

$$= \frac{16}{18} \times \frac{1}{2.59} = 0.3432$$

Efficiency of RBD as compared to CRD

$$= \frac{\text{Amount of information in RBD}}{\text{Amount of information in CRD}} = \frac{0.3432}{0.0939} \times 100 = 365.49\%$$

Gain in efficiency = 365.49 percent

Latin Square Design

Another type of experimental design, which is more restrictive then RBD is Latin Square Design (LSD). Here the number of plots is the square of the no. of treatments and are normally used in experiments to remove the heterogeneity of experimental material in two directions. For example in field experiments when the fertility variation is in two perpendicular directions or there is diagonal trend of fertility or the direction of fertility variation is not known. In animal science or other experiments when there are two source of variation

which need two way blocking e.g. for the data on milk yield the breed and the lactation number are two source of variation. Like wise in experiment on increase in weights of pigs, the initial weights and the breed of pigs are the two source of variation.

For testing v treatments v^2 units are divided into v rows and v columns and the treatments are allotted in such a way that each treatment occurs exactly once in each row and once in each column. This arrangement is called latin square. The term v is called order of the latin square. There is two way blocking one as row blocks and other as column blocks such that each row as well each column is a complete replicate.

Example:

To randomize a Latin square we start with a basic Latin square. Such arrangements are readily available in the Tables for Statisticians and Biometricians (Fisher and Yates, 1948, 1974) for a given order. For example we have 5 x 5 Latin square.

Columns

Rows	1	2	3	4	5
1	A	B	C	D	E
2	B	C	D	E	A
3	C	D	E	A	B
4	D	E	A	B	C
5	E	A	B	C	D

Now the rows are randomized

Columns

Rows	1	2	3	4	5
1	A	B	C	D	E
2	B	C	D	E	A
5	E	A	B	C	D
4	D	E	A	B	C
3	C	D	E	A	B

Next columns of the row randomized square are randomized.

Columns

Rows	5	2	3	1	4
1	E	B	C	A	D
2	A	C	D	B	E
5	D	A	B	E	C
4	C	E	A	D	B
3	B	D	E	C	A

As a result of row and column randomization, but not the randomization of individual units, the whole arrangement remains a latin square.

Advantages

(1) With a two way grouping this design controls more of the variation than CRD and RBD. The two way elimination of variation often results in small error mean square

(2) The analysis is simple

(3) The analysis remains relatively simple even with missing data. Analytical procedures are available for omitting one or more treatment, rows or columns.

(4) Latin square design have a wide variety of application for experimental work. They are used in industry, laboratory, field, green houses, educational, medical, marketing and sociological experiments.

Disadvantage

(1) The number of treatment is limited to the umber of rows and columns. In general 4 x 4 to 12 x 12 Latin squares should be used in practice. For the latin square less than 4 x 4 order, the error d.f. would be too small and the number of experimental units would be too large for the designs beyond 12 x 12 size.

(2) Repeated Latin square of the size 2 x 2 or 3 x 3 are found useful in certain situations to keep the error degree of

freedom reasonably adequate.

Analysis of Latin Square Designs

In Latin square design there are three factors. The data collected from this design are therefore analyzed as a three way classified data. Here the model is

$$Y_{ijk} = \mu + r_i + c_j + t_k + e_{ijk}. \qquad i, j, k = 1, 2, \ldots\ldots v$$

Where y_{ijk} is the observation pertaining to i^{th} row, j^{th} column and k^{th} treatment.

μ, r_i, c_j, t_k (i,j, k = 1 .... v) are fixed effects denoting in order the general mean, the row, the column, and the treatment effects. The e_{ijk}'s are error component assumed to be independently and normally distributed with zero mean and constant variance σ^2.

The analysis is conducted by following a similar procedure as described in RBD.

Let the data be arranged first in a row x column table such that y_{ij} denotes the observation of $(i,j)^{th}$ cell of table.

	Columns				
	1	2	....j.....	v	Total
1					R_1
2					
Rows					
i			Y_{ij}		R_i
.					
.					
v					R_v
Total	C_1		C_j	C_v	G

Let $R_i = \sum_j y_{ij}$ = i^{th} row total (i = 1, 2, ... v)

$C_j = \sum_i y_{ij}$ = j^{th} column total (j = 1, 2.. v)

T_k = Sum of those observations which comes from k^{th} treatment

(k = 1, 2... v)

$$G = \sum_{i=1}^{v} R_i = \sum_{j=1}^{v} C_j = \sum_{k=1}^{v} T_k$$

$$\text{Correction Factor CF} = \frac{G^2}{v^2}$$

$$\text{Total sum of squares (TOT SS)} = \sum_{i,j} Y_{ij}^2 - CF$$

$$\text{Treatment sum of squares (SST)} = \frac{\Sigma T_k^{\,2}}{v} - \text{C.F.}$$

$$\text{Rows Sum of squares (SSR)} = \frac{\Sigma R_i^{\,2}}{v} - \text{C.F.}$$

$$\text{Column sum of squares (SSC)} = \sum_{j} \frac{C_j^2}{v} - \text{C.F.}$$

Analysis of variance of v x v latin square design

Source of variation	**d.f.**	**SS**	**MS**	**F**
Rows	v-1	$\frac{\Sigma R_i^{\,2}}{v} - CF = SSR$	$\frac{MSR}{v-1} = MSR$	$\frac{MSR}{MSE} \sim F_{(v-1,\ \text{error d.f})}$
Columns	v-1	$\frac{\Sigma C_i^{\,2}}{v} - CF = SSC$	$\frac{SSC}{v-1} = MSC$	$\frac{MSR}{MSE} \sim F_{(v-1,\ \text{error d.f})}$
Treatment	v-1	$\frac{\Sigma T_k^{\,2}}{v} - CF = SST$	$\frac{SST}{v-1} = MST$	$\frac{MSR}{MSE} \sim F_{(v-1,\ \text{error d.f})}$
Error	(v-1) (v-2)	By subtraction (SSE)	$\frac{SSE}{(v-1)(v-2)} = MSE$	
Total	v^2-1	$\sum_{ij} Y^2_{ijk}$ - **CF**		

The hypothesis of equal treatment effects is tested by F test. F is

the ratio of treatment mean squares to error mean square. If F is not significant treatment effect do not differ significant among themselves. If F is significant further studies to test the significance of any two treatment means can be made exactly in the same way as in RBD

$$SE_d = \sqrt{\frac{2MSE}{v}}$$

CD at α level significance = SE_d x t [(α, (v-1) (v-2)]

Efficiency

Relative efficiency of Latin square design in comparison to with RBD and CRD.

Let the sketch of ANOVA of LSD be

Source	d.f.	MSS
Rows	n_r	E_r
Columns	n_c	E_c
Treatment	n_t	E_t
Error	n_e	E_e

Relative Efficiency of LSD in comparison with RBD

(1) If row classification had not been made i.e. it the columns were treated as blocks, then estimate of error means square in RBD

$$s_2^2 = \frac{n_r E_r + (n_t + n_e)E_e}{n_r + n_t + n_e}$$

Taking LSD as design D_1 and RBD as D_2 we can find the Relative Efficiency using Fisher's formula

D_1 (LSD)	D_2 (BDR)
$n_1 = n_e$	$n_2 = (n_r + n_e)$
$s_1^2 = E_e$	s_2^2 = as obtained above.

The RE of D_1 (LSD) as compared to D_2 (RBD) is

$$RE = \frac{(n_1+1)(n_2+3)s_2^2}{(n_2+1)(n_1+3)s_1^2} \times 100\%$$ and Gain in efficiency

= (RE -100)

(2) It column classification had not been done i.e. if the rows were treated as block then

Estimate of Error mean square in RBD

$$s_2^2 = \frac{n_c E_c + (n_t + n_e)E_e}{n_c + n_t + n_e}$$

Here

D_1 (LSD)	D_2 (RBD)
$n_1 = n_e$	$n_2 = n_c + n_e$
$s_1^2 = E_e$	s_2^2 = as above.

and using Fisher formula as in the previous case we can find the RE and gain in efficiency of LSD in comparison with RBD.

(3) Relative Efficiency of LSD in comparison with CRD when both rows and columns are merged then

Estimate of Error mean square in CRD

$$s_2^2 = \frac{n_r E_r + n_c E_c + (n_t + n_e)E_e}{n_r + n_c + n_t + n_e}$$

Taking

D_1 (LSD)	D_2 (CRD)
$n_1 = n_e$	$n_2 = n_t + n_c + n_e$
$s_1^2 = E_e$	s_2^2 = as above.

We can find the RE of LSD in comparison with CRD as discussed earlier.

Example

A bacteriologist wanted to determine the effect of four different sources of inoculums, A, B, C, and D and a control, E, on the dry matter yield of a crop. The plots were to be furrow irrigated and the

agronomist suspected that there might be a differential effect of irrigation from one end of the furrow to the other. The field was also bordered by a row of trees and he suspected a differential effect of shading as distance from the tree line increased. He decided to use square plots in Latin square design with rows across the irrigation gradient and columns across the shade gradient. The field plan and total dry matter yields (kg/plot) are given below.

A 32.8	B 32.9	D 31.6	C 33.5	E 25.4
D 36.4	E 29.8	B 32.6	A 33.1	C 34.3
C 35.3	D 36.5	A 37.2	E 27.6	B 34.6
E 34.2	A 37.8	C 36.8	B 38.0	D 35.2
B 33.8	C 38.7	D 32.6	E 37.6	A 35.9

Field plan and dry matter yields (kg/plot) for an inoculum source trial.

Computation

Row-by- column summary table of data from a Latin square design

	Column					
Row	**1**	**2**	**3**	**4**	**5**	**Sum**
1	32.8	32.9	31.6	33.5	25.4	156.2
2	36.4	29.8	32.6	33.1	34.3	166.2
3	35.3	36.5	37.2	27.6	34.6	171.2
4	34.2	37.8	36.8	38.0	35.2	182.0
5	33.8	38.7	32.6	37.6	35.9	178.6
Sum	172.5	175.7	170.8	169.8	165.4	854.2

1. Sums of squares:

$$CF = \frac{G^2}{v^2} = \frac{(854.2)^2}{25} = 29{,}186.31$$

$$SS\ Tot = \sum_i \sum_j Y_{ij}^2 - CF = 29{,}434.40 - 29{,}186.31 = 248.09$$

Treatment Totals and Means of Dry Matter

Treat.	A	B	C	D	E	Sum
Total	176.8	171.9	178.6	177.3	149.6	854.2
Mean	35.36	34.38	35.72	35.46	29.92	

$$SSR = \frac{1}{v}\sum_i R_i^2 - CF = \frac{1}{5}(156.2^2 + ... + 178.6^2) - CF = 84.15$$

$$SSC = \frac{1}{v}\sum_j C_j^2 - CF = \frac{1}{5}(172.5^2 + ... + 165.4^2) - CF = 11.41$$

$$SST = \frac{1}{v}\sum_k T_k^2 - CF = \frac{1}{5}(176.8^2 + ... + 149.6^2) - CF = 117.95$$

$$SSE = SSTOT - SSR - SSC - SST$$

$$= 248.09 - 84.15 - 11.41 - 117.95 = 34.58$$

ANOVA Table

Source	d.f.	SS	MS	F
Rows	4	84.15	21.04	2.43
Columns	4	11.41	2.85	0.99
Treatments	4	117.95	29.49	10.23**
Error	12	34.58	2.88	
Total	24	248.09		

** Significant at the 1% level

Here Tabulated $F_{0.05(4,12)} = 3.26$ $F_{0.01(4,12)} = 5.41$

Since 10.23 > 5.41, treatment differences were highly significant.

Since 2.43 < 5.41, rows were not effective in reducing error.

Since 0.99 < 3.26, columns were not effective in reducing error.

Standard Errors

a. Treatment means:

$$SE_m = \sqrt{\frac{MSE}{v}} = \sqrt{\frac{2.88}{5}} = 0.76$$

b. Difference between two treatment means:

$$SE_d = \sqrt{\frac{2MSE}{v}} = \sqrt{\frac{(2)(2.88)}{5}} = 1.07$$

$$CD = SE_d \times t_{(0.01, 12)} = 1.07 \times 3.05 = 3.27$$

Mean dry matter yield (kg/plot) with Various Sources of Added Incolumn.

Source	A	B	C	D	None	SE	CD at 1%
Mean Yield	35.36	34.38	35.72	35.46	29.92	0.76	3.27

Statistical analysis of data from an experiment designed to measure the effect of different sources of seed inoculums on the dry matter yield of a crop indicates that differences among treatment means were highly significant.

There appeared to be no differential effect of irrigation as well as distance from shade from one end to the other.

❐ ❐ ❐

Chapter – 5

Repeated Latin Squares, Graeco Latin Square, Cross Over and Youden Square Designs

A latin square design is not normally used when the number of treatments is eight or more, because a design of higher order requires too many replications and it may not also be possible to get the required type of experimental units. Again, if the number of treatments is very small, then also there are difficulties for adopting a latin square design. When there are two treatments a latin square design cannot be adopted, because from a 2 X 2 latin square design error variance is not estimable. For a 3 X 3 latin square design the degrees of freedom for error sum of squares is 2 and for a 4 x 4 latin square it is 6. In both these cases the degrees of freedom for error sum of squares are too small. To make the design more effective in such cases, the latin square design may be repeated, that is, instead of taking one latin square, a number of, say, p latin squares each of the same order, is taken for the experiment. The treatments are the same in each square but each has a separate set of units and a separate randomization. The data obtained from such repeated latin squares are analyzed as below.

Each of the p latin squares is first analyzed separately by following the method described in previous chapter. The corresponding sums of squares obtained from the different squares are then added. This gives the pooled row, column, treatment and error sums of squares. The pooled row sum of squares is also called between rows within square sum of squares and similarly for the

other pooled sums of squares.

From each of the p squares, the k treatment totals are obtained, With these totals a p x k table of squares x treatment is prepared.

Let P_i denote the total of all observations in i^{th} square, (i = 1, 2,..., p)

T_s denote the total of observations of s^{th} treatment from all the latin squares, (s = 1, 2, ..., k)

and T_{is} denote the total of all observations of s^{th} treatment in i^{th} square. The square x treatment table is formed with these T_{is} totals. The totals P_i and T_s are the marginal totals of the Square x Treatment Table.

Next, the following sums of squares are obtained:

Correction factor = $(\sum_i P_i^2)/pk^2$

Sum of squares due to squares = $\sum_i \frac{p_i^2}{k^2} - C.F.$

Sum of squares due to treatment = $\sum_s \frac{T_s^2}{kp} - C.F.$

Treatment x square interaction

= (Pooled treatment s.s.) - $\left(\sum_s \frac{T_s^2}{kp} - C.F. \right)$

Total sum of squares = $\sum y_{ijst}^2 - C.F.$

where, y_{ijst} denotes the observation from t^{th} square in its i^{th} row, j^{th} column and under s^{th} treatment. The following partitioning of the degrees of freedom of analysis of variance then follows:

The treatment x square SS may be pooled with error SS if there is no treatment x square interaction. If the squares are at different locations, then under certain hypothesis, it is appropriate to use the treatment x square mean square to test the treatment differences. Otherwise the testing and other procedures are the same as in the ordinary latin square designs.

Partitioning of Degrees of Freedom in the Analysis of Variance of Repeated Latin Squares

Sources of variation	d.f.
Squares	p-1
Between rows within squares (pooled)	p (k-1)
Between columns within squares (pooled)	p(k-1)
Treatments	k-1
Treatment x squares interaction	(p-1) (k-1)
Error (pooled)	p (k- 1) (k-2)
Total	pk^2 - 1

Graeco Latin Square Designs

We have seen that by using latin square designs treatment effects can be estimated by eliminating two sources of variation. This technique can be extended further to eliminate more sources of variation. A graeco latin square is one such design where three sources of variation can be estimated. This design can be defined formally as below.

An orthogonal design by which three sources of variation can be eliminated by using only k^2 units where k is the number of treatments is called a k x k graeco latin square design.

A design is said to be orthogonal when the data obtained from it are orthogonal. When all the levels of a controlled factor occur with each level of any other controlled factor in a design, the data obtained from such a design are always orthogonal.

In order to make the data from a graeco latin square orthogonal, we have to take another controlled factor R at k levels in addition to the two controlled factors P and Q introduced for the latin square design, such that each level of R occurs with each level of P and also of Q only once. If treatments in an ordinary latin square design are taken to be levels of the factor R, then such an arrangement is provided by the latin square design. In that case we have to allocate treatments to the units such that each treatment occurs once with each level of the controlled factors P, Q and R in the k^2 units. For

example, let us replace the five treatments symbols A, B, C, D and E in the example of the latin square design written systematically by the five greek letters α, β, γ, δ and θ respectively. These letters represent the five levels of the factor R. We then get the following arrangement of five levels of each of three factors in 5^2 units:

Level of Q	Levels of P				
	α	β	γ	δ	θ
	β	γ	δ	θ	α
	γ	δ	θ	α	β
	δ	θ	α	β	γ
	θ	α	β	γ	δ

Now, five treatments denoted by A, B, C, D and E are to be allotted to the above 25 units such that each treatment occurs once in each row, each column and with each of the greek letters.

It is not easy to allot the treatments to the units so as to satisfy the above requirements of a graeco latin square design. The problem can, however, be solved easily by exploiting the properties of orthogonal latin squares described below.

Orthogonal Latin Squares

Two latin squares each of the same order, say, r are said to be orthogonal if when one is superimposed on the other each symbol of one falls on each symbol of the other once and only once.

For example, the following two latin squares are orthogonal.

Square I		
A	B	C
B	C	A
C	A	B

Square II		
α	β	γ
γ	α	β
β	γ	α

Orthogonal latin squares are also available in the Tables for Statisticians and Biometricians by Fisher and Yates (1948).

For construction of a graeco latin square design we have to take two orthogonal latin squares each of order k. One of these squares is written by using latin letters and the other by using greek letters. By

superimposing one square over another and by treating the greek letters as the levels of the third controlled factor R, we get the graeco latin square design. The rows and columns of the square denote levels of the other two controlled factors.

For example, the following is a graeco latin square with five treatments.

αA	βE	γD	δC	θB
βB	γA	δE	θD	αC
γC	δB	θA	αE	βD
δD	θC	αB	βA	γE
θE	αD	βC	γB	δA

The randomization procedure of this design is the same as that of the latin square design. The analysis is alsc the same excepting that an extra component of the sum of squares due to the factor R is to be obtained in addition to the usual sums of squares for the latin square design.

ANALYSIS OF VARIANCE TABLE

Source	d.f	SS	MSS	F
Rows	k-1	$\sum \frac{R_i^2}{k} - C.F.$	MSR	
Columns	k-1	$\sum \frac{C_j^2}{k} - C.F.$	MSC	
Symbols	k-1	$\sum \frac{S_k^2}{k} - C.F.$	MSS	
Treatments	k-1	$\sum \frac{T_i^2}{k} - C.F.$	MST	$\frac{MST}{MSE} \sim F_{(t-1,\text{error df})}$
Error	(k-1) (k-3)	(By subtraction)	MSE	
Total	$k^2 - 1$	$\sum\sum y_{ijke}^2 - C.F.$		

$$SE_m = \sqrt{\frac{MSE}{k}}$$

$$SE_d = \sqrt{\frac{2MSE}{k}}$$

$$CD = SE_d \times t_{(\alpha,\ \text{error d.f})}$$

These designs are supposed to be very efficient but it is very difficult to get situations and the required types of experimental units for adopting such designs. This procedure can be extended further to get hyper-graeco latin square designs which can be used to eliminate four or more sources of variation. These designs can be constructed by superimposing three or more mutually orthogonal latin squares. Utility of such designs is, however, limited.

Cross over Designs or Switch over Designs

Another class of designs to remove the heterogeneity of experimental material in two directions are the cross over designs and like latin square designs they are widely used in animal husbandry experiments, where the performance of units is not uniform throughout the period. For example, while studying the effect of feeds in milk yield experiments, the performance in the morning day may be better than that in the evening or performance in the first half of lactation period in general is better than that in the second half of lactation period.

Here total experimental periods is divided into t-factorial periods i.e. equal to number of treatments or feeds to be tested and the treatments are applied to different units in different fractional periods in such a way that

1. Each unit receives all the t-treatments in one or the other period i.e. each unit is a complete replicate.
2. In each fractional period all treatment occur equal number of times which holds when number of unit r is multiple of number of treatments i.e. $r = mr$.

More generally if v treatments are to be tested over v periods, one can use p repetitions of a latin square design (or different Latin squares) juxtaposed if pv experimental subjects are available. If 16

animals are available for experimenting 4 treatments A, B, C, D one can use the following designs.

Periods	Animals															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	A	B	C	D	A	C	D	B	A	D	B	C	A	B	C	D
2	B	A	D	C	B	D	C	A	B	C	A	D	D	A	B	C
3	C	D	A	B	C	A	B	D	C	B	D	A	C	D	A	B
4	D	C	B	A	D	B	A	C	D	A	C	B	A	C	D	A

Here each treatment occurs four times in each of the fractional period. Since each unit crosses or switches over or changes to different treatments in different fractional periods, so the design is called switch over or cross over or change over design.

Analysis

Let y_{ijk} is the observation in i^{th} unit in j-th period having k-th treatment. i= 1, 2, ..., r, j, k = 1, 2 ..... t. Arrange the data in the following table.

Periods	Units	Total
	1 2 ..... i .....r	
p_1	$y_{ij(k)}$	P_1
p_2		P_2
.		.
.		.
p_j		P_j
.		.
.		.
p_t		P_t
Totals	R_1 R_2 ..... R_i .....R_r	

Here $\Sigma R_i = \Sigma P_i = \Sigma T_k = \Sigma\Sigma y_{ij(k)} = G$

C.F. = G^2/rt

ANOVA TABLE

Source	d.f	SS	MSS	F
Replications	r-1	$\sum \frac{R_i^2}{t} - C.F.$	R	
Periods	t-1	$\sum \frac{P_j^2}{t} - C.F.$	P	
Treatments	t-1	$\sum \frac{T_k^2}{t} - C.F.$	T	$\frac{T}{E} \sim F_{(t-1,\text{error df})}$
Error	(r-1)(t-1)	(By subtraction)	E	
Total	rt - 1	$\sum y_{ij(k)}^2 - C.F.$		

Rest of the analysis is the same as of the basic designs

$$SE_d = \sqrt{\frac{2E}{r}}$$

$$CD = \sqrt{\frac{2E}{r}} \times t_{(\alpha,\ \text{error d.f})}$$

Remarks

Switch over designs are preferred over set of LSD, when units with in each factorial period are homogenous.

Set of L.S.D. can't be used for t=2 as the error df p(t-1)(t-2) = 0 whereas switch over designs are most appropriate under this situation.

Example

Given below is data on milk yield in litres of 4 cows when given two feeds in two fractional periods. Give complete analysis using crossover design.

Table: Cross-over experiment for comparing two feeding treatments

Period	Animal Number				Total
	1	2	3	4	
Early lactation	A(10.5)	B(9.8)	B(6.4)	A(9.5)	36.2
Late lactation	B(8.5)	A(9.2)	A(6.6)	B(6.5)	30.8
Total	19.0	19.0	13.0	16.0	67.0

The calculations are as follows:

$$\text{Correction factor} = \frac{(67.0)^2}{8} = 561.125$$

$$\text{Total SS.} = (10.5)^2 + (9.8)^2 + (6.4)^2 + (9.5)^2 + (8.5)^2 + (9.2)^2 + (6.6)^2 + (6.5)^2 - 561.125 = 19.075$$

$$\text{Animal SS.} = \frac{19.0^2 + 19.0^2 + 13.0^2 + 16.0^2}{2} - 561.125 = 12.375$$

$$\text{Periods S.S.} = \frac{36.2^2 + 30.8^2}{4} - 561.125 = 3.645$$

$$\text{Treatments S.S.} = \frac{(10.5 + 9.2 + 6.6 + 9.5)^2 + (8.5 + 9.8 + 6.4 + 6.5)^2}{4} - 561.125 = 2.645$$

Error S.S. = 19.075 – 12.375 – 3.645 – 2.645 = 0.410

We thus set the ANOVA table

Anova Table

Source	d.f.	S.S.	M.S.	F
Animals	3	12.375	4.125	20.12
Periods	1	3.645	3.645	17.78
Treatments	1	2.645	2.645	12.90
Error	2	0.410	0.205	
Total	7	19.075		

The calculated F value for testing the significance of the treatment effects is 12.90 while $F_{(1,2,0.05)} = 18.51$. Thus we conclude the non-significance of the treatment effects at 0.05 level of significance.

Youden Square Design

A Youden Square Design (YSD) is a design with incomplete columns by means of which two sources of variation can be eliminated. The rows of a YSD form a RBD and the columns form a Balanced Incomplete Block (BIB) design. These are basically symmetrical BIB designs by which the block to block variation can be eliminated. The k units in each block can be thought of occupying k different positions. With the help of YSD the effects of such positions can also be eliminated.

Definition

A YSD is an arrangement of v treatments in a k x v rectangular array such the every symbol occurs exactly once in each row and the columns form a symmetrical BIB design with parameters v=b, r=k, λ.

The k units in each block of a symmetrical BIB design can always be so arranged that each treatment occupies each position once in some block or the other. When a symmetrical BIB design is obtained by developing an initial block such an arrangement is evident.

Following is a YSD with 7 treatments arranged in 3 rows and 7 columns.

1	2	3	4	5	6	7
2	3	4	5	6	7	1
4	5	6	7	1	2	3

Example of 7 treatments arranged in 4 rows and 7 columns

1	2	3	4	5	6	7
3	4	5	6	7	1	2
6	7	1	2	3	4	5
7	1	2	3	4	5	6

The position effects are evidently orthogonal to the treatments as each treatment occurs once in each position. The position effects are similarly orthogonal to the blocks as well. Hence, the analysis of YSD is the same as that of BIB designs excepting that a components of

the sum of squares obtained from the position totals of the observation is subtracted from the error sum of squares of the BIB design.

The Youden squares can also be looked upon as an incomplete v x v latin square with k suitably chosen columns. If a column or a row is omitted from a latin square, the resultant design is always a YSD. All these designs Latin Square Design, cross over designs and Youden square Design falls in the general category of Row and Column designs or designs for two-way elimination of heterogeneity where there can be incomplete rows and columns, complete rows and incomplete columns, incomplete rows and complete columns, complete rows and columns.

Youden squares are useful in experiments requiring to remove the heterogeneity in two directions and where fewer replications of the treatments are available.

Let $Y_{ij(m)} = \mu + \rho_i + \Upsilon_j + \tau_m + e_{ij(m)}$,

where μ is the general mean, ρ_i is the i^{th} row effect Υ_j is the j^{th} column effect, τ_m is the m^{th} treatment effect $e_{ij(m)}$ are random errors assumed to be independently normally distributed with mean zero and variance σ^2. Let $R_1, R_2,\ldots,R_k$ be the k rows totals; $C_1, C_2,\ldots, C_v$ be the v column totals, $T_1, T_2,\ldots, T_v$ be the v treatment totals and G be the grand total. Further let Q_m be the adjusted m^{th} treatment total obtained by subtracting the sum of column means in which m^{th} treatment occurs from Tm (m=1, 2,..., v). It then follows that

$$\hat{\tau} = \frac{k}{\lambda v} Q_m, m = 1, 2, \cdots, v,$$

and the ANOVA table will be as given below:

The significance of the treatment effects can be tested by comparing the F value with $F_{\alpha;v-1,vk-2v-k+2}$. The calculation of treatment S.S. is identical to the one for a BIB design and the other calculations are routine exercises.

C.D. for comparing any two treatment effects is

$$(t_{\alpha,vk-2v-k+2})(\sqrt{2kMS_e/\lambda v})$$

ANOVA Table for an Youden Square Design

Source	d.f.	S.S.	M.S.	F
Rows	k - 1	$\sum_{i=1}^{k} \frac{R_i^2}{v} - \frac{G^2}{vr}$	MSR	
Columns	v - 1	$\sum_{j=1}^{v} \frac{C_j^2}{k} - \frac{G^2}{vr}$	MSC	
Treatments (eliminating rows and columns	v - 1	$\sum_{m=1}^{v} \hat{\tau}_m Q_m = \sum_{m=1}^{v} \frac{k}{\lambda v} Q^2{}_m$	MST	MST/MSE ~ $F_{(v-1,\ \text{error df})}$
Error	vk-2v-k+2	By subtraction	MSE	
Total	vk – 1	$\sum_{i,j}^{v} y^2{}_{ij} - \frac{G^2}{vk}$		

* Error S.S. = Total S.S. – Rows S.S. – Column S.S. – Treatment (eliminating rows and columns) S.S.

❐ ❐ ❐

Chapter – 6

Sampling in Field Experiments

In agricultural field experiments, the size of the plot is selected in order to achieve a prescribed degree of precision for measurement of the character of primary interest. We then measure the character under study on the whole of the experimental unit i.e. plot. Because of the nature of character of primary interest like yield, the plot size required is often larger than that needed to measure other characters. In order to save expenses and time the measurements can be made by sampling a fraction of the whole plot. For example, for plant height, the measurement can be made only from say 10 of the 200 plants in the plot, for tiller number, count only $1m^2$ of the $15m^2$ plot, for leaf area, measure only 20 of the approximately 2000 leaves in the plot. For such cases like plant height, leaf area etc. it may not be always possible or desirable to get the plot wise measurement. Here we resort to sampling in each plot and obtain the measurements on a certain number of sampling units in each plot and subject the data for statistical analysis.

Standard analysis of variance requires that there is single observation per character per experimental unit, which is not directly applicable to multi observation data. It can only be applied to the average of all samples from a plot, or to the average of all measurement made over time for each plot. We here discuss appropriate procedures for directly analyzing such data.

Analysis

Suppose an experiment is conducted with t-treatments replicated r times and let there be s observation made in each plot. We assume the following linear additive model for the block design

$$Y_{ijk} = \mu + t_i + B_j + e_{ij} + n_{ijk}$$

where Y_{ijk} is the observation on the k^{th} sampling unit for the i^{th} treatment in the j^{th} replicate (i= 1, 2, ...t; j = 1, 3, ...r ; k = 1, 2.....s), μ is the general mean effect, t_i is the effect of i^{th} treatment, B_j is the effect of j^{th} replication, e_{ij} is the plot error distributed as N (0, σ_e^2), n_{ijk} is the sampling error distributed as N(0, σ_s^2).

Table-1: Format for the analysis of variance of Data from Plot Sampling in a CRD with t Treatments, r Replication, and s Sampling Units per plot.

Source of variation	Degree of Freedom	Sum of Squares	Mean squares	Computed F
Treatment	t-1			
Experimental error	t(r-1)			
Sampling error	rt(s-1)			
Total	rts - 1			

Here an additional source of variation can be measured i.e., due to sampling variation, which is commonly known as sampling error. The format for analysis of variance for data from plot sampling of Completely Randomized Design (CRD) and a randomized block design (RBD) with t treatment and r replications; and for a split plot design with a main plot treatment and b subplot treatment and r replication (However, the split plot design shall be discussed later) is given in table 1, 2 and 3.

The computation procedure is illustrated for RBD and split plot

design with the help of an example. Here the only distinct feature of the analysis variance for data from plot sampling is the part involving sampling error.

Table-2: Format for the Analysis of Variance of Data from Plot Sampling in a RBD Design with t Treatments, r Replications and s Sampling Units per Plot

Source of variation	Degree of Freedom	Sum of Squares	Mean squares	Computed F
Replication	r-1			
Treatment	t-1			
Experimental error	(r-1)(t-1)			
Sampling error	rt(s-1)			
Total	rts - 1			

Table-3: Format for the Analysis of Variance of Data from Plot Sampling in a Split-Plot Design with a Main Treatments, b Subplot Treatments, r Replications and s Sampling Units per Plot

Source of variation	Degree of Freedom	Sum of Squares	Mean squares	Computed F
Replication	r-1			
Main plot treatment (A)	a-1			
Error (a)	(r-1) (a-1)			
Subplot treatment (B)	b-1			
A x B	(a-1)(b-1)			
Error (b)	a (r-1) (b-1)			
Sampling error	abr(s-1)			
Total	rabs - 1			

Randomized Block Design

An experiment with nine treatments in four replications was conducted in Randomized Block Design. Data on tiller count, collected from four randomly selected 2 x 2 hills sampling units is give below:

Tiller Count (no./4hills) of Rice tested under nine fertilizer treatments in a RBD experiment with four replications and four sampling units (S_1, S_2, S_3 and S_4)

Treatments	Rep. I				Rep.II				Rep.III				Rep.IV			
Number	S_1	S_2	S_3	S_4	S_1	S_2	S_3	S_4	S_1	S_2	S_3	S_4	S_1	S_2	S_3	S_4
1	31	24	26	20	25	23	26	31	33	25	34	22	38	40	39	27
2	46	44	35	45	60	58	40	44	68	62	60	60	38	55	38	62
3	41	46	60	48	47	38	41	68	50	46	53	52	49	63	55	71
4	42	48	41	52	67	63	74	51	78	58	72	71	55	39	50	60
5	48	66	57	50	48	74	53	48	54	40	47	60	52	50	38	74
6	65	60	55	41	53	49	57	52	72	55	44	53	61	55	44	50
7	55	42	67	51	44	48	74	55	78	50	77	51	70	78	68	78
8	60	58	55	47	42	55	66	62	44	55	62	72	66	75	82	64
9	77	65	77	45	52	45	43	50	66	57	55	59	51	53	48	47

Computations

Construct the replication x treatment table of totals and compute replication total (R), the treatment total (T) and the grand total G.

The Replication x Treatment Table of Totals Computed from Data.

Treatment Number	**Tiller Count Total (RT)**				**Treatment Total (T)**
	Rep.I	**Rep.II**	**Rep.III**	**Rep.IV**	
1	101	105	114	144	464
2	170	202	250	193	815
3	195	194	201	238	828
4	183	255	279	204	921
5	221	223	201	214	859
6	221	211	224	210	866
7	215	221	256	294	986
8	220	225	233	287	965
9	264	190	237	199	890
Rep. total (R)	1790	1826	1995	1983	
Grand total (G)					7594

$$C.F. = G^2/trs = \frac{(7594)^2}{9 \times 4 \times 4} = 400478.00$$

Total ss = $\Sigma x^2 - C.F.$

$= [(31)^2 + (24)^2 + \ldots\ldots (47)^2] - 400478 = 26535$

$$\text{Replication SS} = \frac{\Sigma R^2}{ts} - C.F.$$

$$= \frac{(1790)^2 + (1,826)^2 + (1995)^2 + (1,983)^2}{(9)(4)} - 400478 = 930.027$$

$$\text{Treatment SS} = \frac{\Sigma T^2}{rs} - C.F.$$

$$= \frac{(464)^2 + (815)^2 + \ldots + (890)^2}{(4)(4)} - 400478 = 11815.97$$

$$\text{Experimental errors SS} = \frac{\Sigma (RT)^2}{s} - C.F. - \text{Replication SS} - \text{Treatment SS}$$

$$= \frac{(101)^2 + (105)^2 + (114)^2 + \ldots + (199)^2}{4}$$

$- 400478 - 930.027 - 11815.97 = 4720.972$

Sampling error SS = Total SS – (sum of all other SS)

$= 26535.97 - (930.027 + 11815.9 + 4720.972) = 9069$

The format of analysis of variance is given in table 2.

Analysis or variance (RCB with Data from Plot Sampling) of Data

Source of variation	Degree of Freedom	Sum of Squares	Mean squares	Computed F
Replication	3	930.03	310.01	
Treatment	8	11815.97	1477.00	7.51**
Experimental error	24	4720.97	196.71	2.34**
Sampling error	108	9069.00	83.97	
Total	143	26535.97		

** Significant at 1% level.

The mean square (s_1^2) is first tested against s_2^2 if (i) s_1^2 is significant than treatments are tested against s_1^2 and if (ii) s_1^2 is not significant then treatments are tested against the pooled mean square of s_1^2 and s_2^2. Here s_1^2 is significant, so we test the treatment against s_1^2. F(8, 24) at 1% level of significance is 3.36 which is less than the computed value of F, so treatments differ significantly.

$$S.E_d = \sqrt{\frac{2s_1^2}{rs}}$$

where s_1^2 is the experimental error MS in the analysis of variance. For our example, the standard error of the difference between any pair of treatments is:

$$S.E_d = \sqrt{\frac{2(196.71)}{(4)(4)}} = 4.96$$

CD at 1% = $t_{(.01, 24)} \times S.E_d = 2.79 \times 4.96 = 13.83$

Here the estimates of the sampling error variance and of the experimental error variances are:

$$s_S^2 = s_1^2$$

$$s_E^2 = \frac{s_2^2 - s_1^2}{s}$$

where s_1^2 is the sampling error MS in the analysis of variance. For our example, the two variance estimates and their corresponding cv values are

$$s_s^2 = 83.97$$

$$s_E^2 = \frac{196.71 - 83.97}{4} = 28.18$$

$$cv(S) = \frac{\sqrt{83.97}}{52.73} \times 100 = 17.38\%$$

$$cv(E) = \frac{\sqrt{21.31}}{52.73} \times 100 = 10.067\%$$

Split – Plot Design

An experiment was conducted in a split-plot design involving eight management levels as main-plot treatments and four of nitrogen application as subplot treatments. There are three replications. The data on plant height measured on two single-hill sampling units per plot, are given below.

Height of rice plants measured on two sampling units (S_1 and S_2) per plot, from a Split-Plot Experiment involving eight management levels (M_1, M_2,.....M_8) and four times of Nitrogen application (T_1, T_2, T_3, and T_4) with three replications.

Treatment Combination		Plant Height (cm)					
Time of Application	Management Level	Rep-I		Rep.II		Rep.III	
		S_1	S_2	S_1	S_2	S_1	S_2
T_1	M_1	102.5	108.5	113.2	107.5	110.3	105.9
	M_2	90.5	89.6	115.4	110.8	106.9	104.3
	M_3	94.5	93.6	110.5	115.8	104.9	106.9
	M_4	93.5	91.3	109.7	104.6	110.8	106.9
	M_5	104.8	102.3	108.6	100.9	102.7	96.8
	M_6	95.8	107.9	100.4	110.5	103.8	109.6
	M_7	95.0	93.0	98.0	100.8	104.9	116.9
	M_8	100.5	112.3	119.5	104.6	97.5	100.3
T_2	M_1	110.5	108.4	115.6	106.0	115.9	108.7
	M_2	95.6	97.8	120.5	116.7	101.3	105.0
	M_3	93.5	91.7	107.0	104.6	116.8	123.0
	M_4	100.0	105.0	116.0	117.0	108.0	110.0
	M_5	96.0	98.0	108.8	106.6	118.3	116.9
	M_6	100.3	95.8	109.7	111.5	118.9	122.5
	M_7	99.7	101.8	98.6	99.5	101.5	105.6
	M_8	105.8	106.8	109.9	106.8	108.4	110.3
T_3	M_1	96.5	94.5	108.9	110.5	116.5	118.0
	M_2	94.0	93.0	108.9	110.5	109.3	105.6
	M_3	90.0	91.8	108.9	109.6	110.5	114.3
	M_4	90.0	91.3	106.3	108.5	112.3	114.9
	M_5	102.5	103.5	98.6	99.5	93.4	95.6

	M_6	99.5	98.6	101.5	101.0	111.3	109.8
	M_7	90.3	91.3	96.1	94.2	100.0	98.6
	M_8	108.9	110.0	111.3	113.4	100.0	98.0
T_4	M_1	108.3	110.0	106.8	105.7	118.0	115.6
	M_2	89.6	90.5	116.0	121.5	93.4	101.5
	M_3	93.5	99.6	115.6	117.8	119.2	121.5
	M_4	104.9	107.8	105.6	116.9	121.5	124.6
	M_5	100.0	101.3	116.5	119.6	126.8	127.6
	M_6	95.0	96.0	101.5	102.8	120.3	121.8
	M_7	95.8	97.8	100.8	100.9	111.5	116.5
	M_8	112.8	113.8	95.8	94.3	112.5	114.9

Let the main plot factor is A, subplot factor is B. Denote the levels of factor A by a, the levels of factor B by b, the number of replication by r and the number of sampling units per plot by s.

Step (i) Construct Replication x factor A two way table of totals.

The Replication x Management Level Table of Totals (RA) Computed from Data.

Management Level	**Plant Height Total (RA)**			**Management Total (A)**
	Rep.I	**Rep.II**	**Rep.III**	
M_1	839.3	874.2	908.9	2622.4
M_2	740.6	920.3	827.3	2488.2
M_3	748.2	889.8	917.0	2555.0
M_4	783.8	884.6	909.0	2577.4
M_5	808.4	859.1	878.1	2545.6
M_6	788.9	838.9	918.0	2545.8
M_7	764.7	788.9	855.5	2409.1
M_8	870.9	855.6	841.9	2568.4
Rep. total (R)	6344.8	6911.4	7055.7	
Grand total (G)				20311.9

$$C.F. = \frac{G^2}{rabs} = \frac{(20311.9)^2}{(3)(8)(4)(2)} = 2148819.00$$

Step (ii) Construct factor A x factor B two way table of totals (AB)

The Management Level x Time of Nitrogen Application Table of Totals Computed from Data.

Management Level	Plant Height Total (AB)			
	T_1	T_2	T_3	T_4
M_1	647.9	665.1	645.0	664.4
M_2	617.5	636.9	621.3	612.5
M_3	626.1	636.6	625.1	667.2
M_4	616.8	656.0	623.3	681.3
M_5	616.1	644.6	593.1	691.8
M_6	628.0	658.7	621.7	637.4
M_7	608.6	606.7	570.5	623.3
M_8	634.7	648.0	641.6	644.1
Total (B)	4995.7	5152.6	4941.6	5222.0

Total SS $= \Sigma X^2 - C.F.$

$= [(102.5)^2 + (108.5)^2 + \ldots + (114.9)^2 - 2148819 = 4992.63$

Replication SS $= \dfrac{\Sigma R^2}{abs} - C.F.$

$= \dfrac{(6344.8)^2 + (6911.4)^2 + (7055.7)^2}{(8)(4)(2)} - 2148819 = 4412.692$

A (management level) SS $= \dfrac{\Sigma A^2}{rbs} - C.F.$

$= \dfrac{(2622.4)^2 + (2488.2)^2 + \ldots + (2568.4)^2}{(3)(4)(2)} - 2148819 = 1212.289$

Error (a) SS $= \dfrac{\Sigma(RA)^2}{bs} - C.F. - \text{Replication SS} - \text{A SS}$

$= \dfrac{(839.3)^2 + (874.2)^2 + \ldots + (841.9)^2}{(4)(2)}$

$- 2148819 - 4412.692 - 1212.289 = 3053.181$

B (time of application) SS $= \dfrac{\Sigma B^2}{ras} - C.F.$

$= \dfrac{(4,995.7)^2 + (5152.6)^2 + (4941.6)^2 + (5222.0)^2}{(3)(8)(2)} - 2148819$

= 1076.654

Step (iii) Construct Replication x factor A x factor B three way table of totals (RAB)

Management Level	Time of Application	Plant Height Total (RAB)		
		Rep. I	Rep. II	Rep. III
M_1	T_1	211.0	220.7	216.2
	T_2	218.9	221.6	224.6
	T_3	191.1	219.4	234.5
	T_4	218.3	212.5	233.6
M_2	T_1	180.1	226.2	211.2
	T_2	193.4	237.2	206.3
	T_3	187.0	219.4	214.9
	T_4	180.1	237.5	194.9
M_3	T_1	188.1	226.3	211.7
	T_2	185.2	211.6	239.8
	T_3	181.8	218.5	224.8
	T_4	193.1	233.4	240.7
M_4	T_1	184.8	214.3	217.7
	T_2	205.0	233.0	218.0
	T_3	181.3	214.8	227.2
	T_4	212.7	222.5	246.1
M_5	T_1	207.1	209.5	199.5
	T_2	194.0	215.4	235.2
	T_3	206.0	198.1	189.0
	T_4	201.3	236.1	254.4
M_6	T_1	203.7	210.9	213.4
	T_2	196.1	221.2	241.4
	T_3	198.1	202.5	221.1
	T_4	191.0	204.3	242.1
M_7	T_1	188.0	198.8	221.8
	T_2	201.5	198.1	207.1
	T_3	181.6	190.3	198.6
	T_4	193.6	201.7	228.0
M_8	T_1	212.8	224.1	197.8
	T_2	212.6	216.7	218.7
	T_3	218.9	224.7	198.0
	T_4	226.6	190.1	227.4

$$A \times B\ SS = \frac{\Sigma(AB)^2}{ras} - C.F. - B\ SS - A\ SS$$

$$= \frac{(647.9)^2 + (665.1)^2 + ... + (644.1)^2}{(3)(2)}$$

- 2148819-1076.654-1212.289 = 981.5536

$$\text{Error (b) SS} = \frac{\Sigma(RAB)^2}{s} \text{-C.F. – Replication SS – A SS}$$

- Error (a) SS – B SS – A x B SS

$$= \frac{(211.0)^2 + (220.7)^2 + ... + (227.4)^2}{2}$$

- 2148819 – 4212.692 – 1212.289

- 3053.181 – 1076.654 – 981.5536 = 3314.981

Sampling error SS = Total SS – (Sum of all other SS)

= 14992.63 – (4412.692 + 1212.289 + 3053.181

+ 1076.654 + 981.5536 + 3314.981) = 941.285

Analysis of Variance (Split-Plot Design with Data from Plot Sampling)

Source	Degree of Freedom	Sum of Squares	Mean Square	Computed F
Replication	2	4412.692	2206.346	
Management level (A)	7	1212.289	173.1841	< 1
Error (a)	14	3053.181	218.08 (E_a)	
Time of application (B)	3	1076.654	358.8848	5.20**
A x B	21	981.5536	46.74065	<1
Error (b)	48	3314.981	69.06 (E_b)	
Sampling error	96	941.285	9.80 (E_s)	
Total	191	14992.63		

** = significant at 1% level

The result indicates that only the effect of the time of nitrogen

application is significant.

For pair comparison, compute the standard error of the mean difference following the appropriate formula but with one modification – multiply each divisor by s, the sample size in each plot. For example, to compare two subplot treatments at the same main-plot treatment, the standard error of the mean difference is:

$$S.E_d = \sqrt{\frac{2E_b}{rs}} = \sqrt{\frac{2x\ 69.06}{3x2}} = 4.80$$

Where E_b is the error (b) MS from the analysis of variance. And, to compare two subplot treatments (averaged over all main-plot treatments) the standard error of the mean difference is:

$$S.E_d = \sqrt{\frac{2E_b}{ras}} = \sqrt{\frac{2x\ 69.06}{3x8x2}} = 1.70$$

Now the C.D. values can be calculated and significance of treatment means can be tested.

Compute the estimates of two variance components: the experimental error associated with the smallest experimental unit [i.e. error (b) in this case] and the sampling error as:

$$s_E^2 = \frac{E_b - E_s}{s} = \frac{69.06 - 9.80}{2} = 29.63$$

$$s_S^2 = E_s = 9.80$$

where E_s is the sampling error MS, E_b is the experimental error MS, and s is the number of sampling units per plot.

$$cv(E) = \frac{\sqrt{29.63}}{105.79} x100 = 5.14\ \%$$

$$cv(S) = \frac{\sqrt{9.80}}{105.79} x100 = 2.96\ \%$$

The results indicate a relatively small sampling error compared to the experimental error.

❐ ❐ ❐

Chapter – 7

Missing Plot Technique

A missing data situation occurs whenever a valid observation is not available for any one of the experimental units. Poor germination, physical damage during crop culture, damage due to birds and rodents and pest damage are common causes of destruction of experimental units. When ever there is a missing plot, the experiment cannot be analysed in the usual way, except designs like completely randomized design, where we can analyse the data even with unequal replications. Experiments are usually planned with equal number of observation per cell, which facilitates the analysis. The designs CRD, RBD, Latin squares are called orthogonal designs, here the block contrasts, treatment contrast and error contrast are orthogonal to each other and therefore additive. The effects of the treatments, blocks etc. can be estimated and tested with out any entanglement (i.e. separated). Such a situation in RBD and Latin square designs does not remain so if there is a missing plot. The data becomes non orthogonal and the same can not be analyzed in the usual way of analysis of variance. Forced with such situation, the experimenter would still like to obtain all the pieces of information possible from the remaining observations.

In response to such situations, several analytical procedures for handling such experiments have been developed, which are.

(i) Estimating the missing value(s) using principle of Least Square i.e. minimizing the error sum of squares.

(ii) Method of iteration.

(iii) Method of fitting constants and,

(iv) Analysis of data with missing observations by the technique of analysis of covariance.

Here we shall describe the first two methods of analysis of data with missing observations:

Analysis by estimating the missing value

This method is due to Yates. It consists of estimating the missing observations by the value which makes the error sum of squares minimum. In order to get such estimates, first unknown values are substituted for missing observations. The error sum of squares with the designs is then expressed as a function of this unknown. Next, by differentiating the error sum of squares with respect to the unknown and equating differentiates to zero, as many linear equations in the unknown as the number of the missing observations are obtained. Solutions of these equations provide the estimated values of the missing observations.

The data are then completed by substituting these estimated values and then analysed by the usual techniques appropriate for the design.

Here it is emphasized that an estimate of the missing data obtained through this technique does not supply any additional information to the incomplete set of data, once the data is lost, no amount of statistical manipulation can retrieve it. What the procedure attempts to do is to allow the experimenter to compute the analysis of variance in the usual manner (i.e. as if the data were complete) without restoring to the more complex procedure needed for incomplete data sets.

From this analysis, the correct error sum of squares is obtained but not the treatment sum of squares. In such case, the correct treatment sum of squares is obtained by subtracting the correct error sum of squares from "within block sum of squares" obtained from the

incomplete data by analyzing as CRD, ignoring the treatments. The degrees of freedom of the error sum of squares and the total sum of squares is reduced by the number of missing observations.

One Missing observation in RBD:

Let X be the value of the missing observation

r be the number of replications

t be the number of treatments

B_i' be the total of available observation in the i^{th} block (with the missing observations)

T_j' be the total of available observations of the j^{th} treatment (for which the observation is missing)

G' be the grand total of all the available observations.

Error S.S. = Total S.S. – Replication S.S. – Treatment S.S.

$$= \sum_i \sum_j y_{ij}^2 + X^2 - \left(\frac{1}{t} \sum_{i \neq 1} B_i^2 + \frac{(B_i + X)^2}{t} \right)$$

$$- \left(\frac{1}{r} \sum_{i \neq 1} T_j^2 + \frac{(T_j' + X)^2}{r} \right) + \frac{(G'+X)^2}{rt}$$

This expression is differentiated with respect to X and the differential is equated to zero. The solution of the resulting equation gives the estimate of X, the missing observation, when the error variance is minimum.

The estimated value of x is given by

$$\hat{X} = \frac{rB_i' + tT_j' - G'}{(r-1)(t-1)}$$

The analysis is carried out by substituting the estimated value of X with the following changes:

(i) One d.f. is to be subtracted from the d.f. corresponding to error sum of squares and total sum of squares.

(ii) Adjusting the treatment sum of squares and total sum of squares by subtracting from it a quantity.

$$\frac{(B_i' + tT_j' - G')^2}{t(t-1)(r-1)^2}$$

Standard error for testing the significance of the difference between treatment means.

(a) S.E. of difference of two treatment means not involving missing value

$$SE_d = \sqrt{\frac{2S_e^2}{r}}$$

(b) S.E. of difference of two treatment means one of which involves the missing value.

$$SE_d = \sqrt{S_e^2\left[\frac{2}{r} + \left(\frac{t}{r(t-1)(r-1)}\right)\right]}$$

Example - 1

To find out the best source of nitrogen at 60 kg/ha, an experiment was conducted on paddy in RBD with 5 sources of nitrogen in 4 blocks at Paddy breeding centre, Coimbatore, Tamil Nadu. The yield data for different treatments are given as

Yield of grain in kg/plot

Blocks (replication)	Ammonium Sulphate	Ammonium Chloride	Urea	Chilean Nitrate	Ammonium Sulphate Nitrate	Total
	S_1	S_2	S_3	S_4	S_5	
I	25.4	32.5	37.5	22.5	20.5	138.4
II	17.3	-	25.4	14.7	21.5	78.9
III	22.4	28.4	30.1	23.5	23.5	127.9
IV	30.5	33.4	34.5	22.4	28.5	149.3
TOTAL	95.6	94.3	127.5	83.1	94.0	494.5

The observation relating to application of ammonium chloride in

the second block is missing. Analyse the data.

Calculations

$r = 4,\ t = 5,\ B_i' = 78.9,\ T_j' = 94.3,\ G' = 494.5$

(i) The estimate of missing value, X, is obtained as

$$\hat{X} = \frac{rB_i' + tT_j' - G'}{(r-1)(t-1)} = \frac{4x78.5 + 5x94.3 - 494.5}{3x4} = 24.4$$

It is to be noted that $\hat{X}$ is not the value that would have been obtained if the value of S_2 was not missing in second block. It is the value which, when substituted for missing observation, will enable us to obtain through the usual method of analysis an unbiased estimate of error variance.

We now insert the estimated missing value and carry out the analysis of variance according to the usual procedure for an R.B.D. except for subtracting 1 d.f. from total S.S. as well as from the d.f. for error S.S.

(ii) Calculation of Sum of Squares

G.T. = 494.5 + 24.4 = 518.9

$$C.F. = \frac{(G.T.)^2}{rt} = \frac{(518.9)^2}{20} = 13462.86$$

$$\text{Total S.S.} = \sum_i \sum_j Y_{ij}^2 - C.F. = 14121.21 - 13462.86 = 658.35$$

$$\text{Block S.S.} = \sum_i \frac{B_i^2}{t} - C.F. = 13694.87 - 13462.86 = 232.01$$

$$\text{Treatment S.S.} = \sum_j \frac{T_j^2}{r} - C.F.$$

$$= 13806.73 - 13462.86 = 343.87$$

Error S.S. = Total S.S. – Block S.S. – Treatment S.S.

= 658.35 - 232.01 – 343.87 = 82.47

While the error mean square is an unbiased estimate of the error variance, the treatment S.S. is an over estimate and has to be corrected by subtracting from it a quantity.

$$\frac{(B_i' + tT_j' - G')^2}{t(t-1)(r-1)^2} = \frac{(78.9 + 5x94.3 - 494.5)^2}{5x4x9} = 17.36$$

Therefore, corrected treatment S.S. = 343.87 – 17.36 = 326.51

Similarly corrected total S.S. = 658.35 – 17.36 = 640.99

(iii) Analysis of variance

ANOVA

Source	d.f.	S.S.	M.F.	F
Blocks	3	232.01	77.34	10.31**
Treatment (Unadj)	4	343.87	85.97	11.46**
Treatment (Adj)	4	326.51	81.63	10.88**
Error	11	82.47	7.50	-
Total	18	640.99	35.61	

** indicates significance at 1 percent level

(iv) Standard Errors

(a) S.E. the difference between two treatment means not involving the missing value.

$$SE_d = \sqrt{\frac{2S_e^2}{r}} = \sqrt{\frac{2x7.50}{4}} = 1.936 \text{kg/plot}$$

(b) S.E. the difference between two treatment means one of which is a missing value.

$$SE_d = \sqrt{S_e^2\left[\frac{2}{r} + \left(\frac{1}{r(r-1)(t-1)}\right)\right]} = \sqrt{7.50\left[\frac{2}{4} + \frac{1}{4x3x4}\right]} = 2.13 \text{kg/plot}$$

Two missing values in R.B.D.

We have discussed earlier one method in that analysis of variance for any number of missing observations can be done by

expressing the error sum of squares as function of the unknown and then by differentiating with respect to the unknowns and equating the resulting equations to zero, we get as many equations as the number of unknowns. By solving these equations we get the estimates of the missing values. By proceeding in the like manner, we can analyse the data.

Besides this there is a method known as iterative method (Yates 1933).

Iterative Method

The iterative method of analysis is used when more than one observation is missing. Suppose values of two observations are missing in R.B.D. These missing values can be (i) in the same block (ii) in different blocks and belonging to different treatments. The estimates of the missing values can be obtained by the method of iteration or successive approximations suggested by Yates. The missing values are denoted by X and Y and X_α and Y_α, $\alpha = 1, 2. \ldots p$, denote the values of X and Y in α^{th} iteration process which terminates in p steps. The method consists of substituting an approximate value, Y_1 (usually the mean value of the available observations), for one of the missing observations and the other value is computed as in case of a single missing observations, by the formula:

$$\hat{X} = \frac{rB_i' + tT_j' - G'}{(r-1)(t-1)}$$

The value X_1 is now substituted for X and the second iterated value of Y namely Y_2 is computed using the formula given above. Using this value of Y_2, the value of X namely X_2, is computed again using the same formula. The Process is repeated till each of the values of X and Y obtained from two consecuting iterations are nearly the same. The analysis is then done in the usual way except that from the total SS and the error SS, 2 d.f. are subtracted.

The error mean square so obtained is an unbiased estimate of the variance. The adjusted SS for treatments are calculated as

follows:

Ignoring the treatments and without substituting for the missing observation, we obtain the between and within blocks SS. Error Sum of Squares is then subtracted from this within blocks SS to provide the adjusted treatments SS. To obtain a standard error for the comparison of two treatment means for A and B both having missing values, we assign an effective number of replicates to each treatment say r_1 and r_2. Which are calculated by the following rules:

(i) Any replicate of treatment A is counted as 1 when both the treatment A and B are present in the replicate.

(ii) Counted as (t-2)/(t-1) when A is present and B is not present.

(iii) Counted as zero when A itself is missing

Then the standard Error of difference between treatment means for A and B is

$$SE_{(d)} = \sqrt{MSE(1/r_1 + 1/r_2)}.$$

and CD = $SE_{(d)}$ x $t\alpha$ at error d.f.

Example 2

In the example discussed in example 1, assume that the yield of S_4 in third block is also missing. Analyse the data.

(i) We prepare the following two-way table of treatment and blocks putting X and Y for the missing values:

Yield in kg/plot

Replications	Treatments					Total
	1	2	3	4	5	
I	25.4	32.5	37.5	22.5	20.5	138.4
II	17.3	Y	25.4	14.7	21.5	78.9
III	22.4	28.4	30.1	X	23.5	104.4
IV	30.5	33.4	34.5	22.4	28.5	149.3
TOTAL	95.6	94.3	127.5	59.6	94.0	471.0

(ii) Let Y_1 = 31.4 (the mean value over replication) for the value of Y, thus

$$X_1 = \frac{4 \times 104.4 + 5 \times 59.6 - 502.4}{3 \times 4} = 17.77$$

$$\text{Again } Y_2 = \frac{4x78.9 + 5 \times 94.3 - 488.77}{3 \times 4} = 24.86$$

$$\dot{X}_2 = \frac{4 \times 104.4 + 5 \times 59.6 - 495.86}{3 \times 4} = 18.31$$

$$Y_3 = \frac{4 \times 78.9 + 5 \times 94.3 - 489.31}{3 \times 4} = 24.82$$

$$X_3 = \frac{4 \times 104.4 + 5 \times 59.60 - 495.82}{3 \times 4} = 18.31$$

$$Y_4 = \frac{4 \times 78.9 + 5 \times 94.3 - 489.31}{3 \times 4} = 24.82$$

$$X_4 = \frac{4 \times 104.4 + 5 \times 59.6 - 495.82}{3 \times 4} = 18.31$$

Since the pairs (X_3, Y_3) and (X_4, Y_4) are same i.e. *(18.31, 24.82)*, we now stop further. These are the estimated values of X and Y. we insert these values in the table and analyse the completed data. Here, since there are two missing observations, we have to subtract two d.f. each from the d.f. for the total S.S. and error S.S.

Calculation of Sum of Squares:

$$\text{C.F.} = \frac{(\text{G.T.})^2}{rt} = \frac{(514.1)^2}{4 \times 5} = 13214.94$$

$$\text{Total S.S.} = \sum_i \sum_j y_{ij}^2 - \text{C.F.}$$

$$= 13923.53 - 13214.94 = 708.59$$

$$\text{Replication SS} = \frac{1}{t}\sum B_i^2 - \text{C.F.}$$

$$= 13450.81 - 13214.94 = 235.87$$

$$\text{Treatment} = \frac{1}{r}\sum_i T_j^2 - \text{C.F.}$$

$$= 13621.21 - 13214.94 = 406.27$$

Error S.S. = Total SS – Treatment S.S. – Replication S.S.

$$= 708.59 - 406.27 - 235.87 = 66.45$$

Calculation of adjusted Treatment S.S.

We now have to analyse data of the original table where there are two missing values (i.e. analysis for the available information)

$$\text{C.F.} = \frac{(\text{G.T.})^2}{18} = \frac{(471.0)^2}{18} = 12324.50$$

$$\text{Total S.S.} = \sum_i \sum_j y_{ij}^2 - \text{C.F.}$$

$$= 12973.60 - 12324.50 = 649.10$$

$$\text{Replication S.S.} = \sum_i \frac{B_i^2}{t_i} - \text{C.F.}$$

Where t_i is the number of treatments in the i^{th} block

$$= \frac{(138.4)^2}{5} + \frac{(78.9)^2}{4} + \frac{(104.4)^2}{4} + \frac{(149.3)^2}{5} - 12324.50$$

$$= 245.65$$

Within Replication S.S = Total S.S.- Replication S.S.

$$= 649.10 - 245.65 = 403.45$$

The within replications S.S. contains the correct treatments S.S. + correct error S.S. Therefore by subtracting form it the error S.S. which we have earlier calculated we get the correct treatment S.S.

Therefore, correct treatment S.S. = 403.45-66.45 = 337.00

ANOVA

Source	d.f.	S.S.	M.S.	F
Replication	3	245.65	81.88	
Treatments (Adjusted)	4	337.00	84.25	12.68**
Error	10	66.45	6.645	
Total	17	649.10		

Standard Errors

(a) S.E. of difference of two treatment means one of which involves a missing value.

$$SE_d = \sqrt{\left[\frac{2}{r} + \frac{t(t-1)(r-1)}{r[(r-1)^2(t-1)^2 - 1]}\right] S_e^2}$$

$$= \sqrt{\left[\frac{2}{4} + \frac{5 \times 4 \times 3}{4[(9 \times 16) - 1]}\right] 6.645} = 2.00 \text{ kg/plot}$$

(b) S.E. of difference between the means of the treatments containing estimates of the missing value is

$$SE_d = \sqrt{\left[\frac{2}{r} + \frac{2t}{[r(r-1)(t-1) - 1]}\right] S_e^2}$$

$$= \sqrt{\left[\frac{2}{4} + \frac{2 \times 5}{4 \times 11}\right] 6.645} = 2.198 \text{ kg/plot}$$

(c) S.E. of difference between the means none of which involves a missing value

$$SE_d = \sqrt{\frac{2S_e^2}{r}} = \sqrt{\frac{2 \times 6.645}{4}} = 1.823 \text{ kg/plot}$$

One missing observation in Latin Square design:

The procedure is to first obtain the estimates of missing values X, by the formula

$$\hat{X} = \frac{t(R'+C'+T') - 2G'}{(t-1)(t-2)}$$

where

t is the number of treatments

R' is the total of available observation in the row with missing value.

C' is the total of available observations in the column with missing value.

T' is the total of available observation for the treatment with missing value.

G' is the grand total of all the available observations.

Then the estimated missing value is inserted in the table and the analysis of variance is carried out as per the usual procedure of latin square design but there again we have to subtract 1 d.f. each from the d.f. for total S.S. and the error S.S.

Example

In order to find the effect of mineral treatments on Paddy, an experiment in latin square design was laid down at Agri. Res. Station, Manoda (Mysore) during 1990. The details of the experiment are given below:

	Treatments	**Dose**
1.	Zinc Sulphate	5.60 kg/ha
2.	Zinc Sulphate	11.21 kg/ha
3.	Borax Sulphate	22.42 kg/ha
4.	Copper Sulpahte	5.60 kg/ha
5.	Copper Sulphate	11.21 kg/ha
6.	Control	

Plot Size: 1/100

Yield of grain in kg/plot

Row	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
1	(6) 4.08	(5) 3.63	(4) 3.86	(3) 2.72	(2) 2.72	(1) 5.44	
2	(5) 5.44	(4) 3.63	(1) 2.72	(6) -	(3) 2.49	(2) 6.75	21.03 (R')
3	(4) 5.61	(1) 1.90	(5) 2.81	(2) 3.01	(6) 2.61	(3) 4.25	
4	(3) 5.27	(6) 4.25	(2) 3.12	(5) 4.82	(1) 4.82	(4) 3.06	
5	(2) 2.07	(3) 2.95	(6) 2.49	(1) 2.75	(4) 3.35	(5) 2.38	
6	(1) 4.48	(2) 2.52	(3) 2.55	(4) 2.69	(5) 1.62	(6) 3.18	
Total				15.99(C')			

Note: Figures in bracket indicate the treatment number. Analyse the data and draw conclusion.

Calculation:

(i) We first estimate the missing value, X, as

$$\hat{X} = \frac{t(R'+C'+T')-2G'}{(t-1)(t-2)}$$

$$= \frac{6(21.03+15.99+16.61)-2 \times 122.04}{5 \times 4} = 3.89$$

Now, substitute this value of $\hat{X}$ in the table and proceed for analysis

(ii) Calculation of various sum of squares

$$C.F. = \frac{(G.T.)^2}{t^2} = \frac{(125.93)^2}{36} = 440.5$$

Total S.S. = 491.02 – 440.51 = 50.51

S.S. due to rows = 453.47 – 440.51 = 12.96

S.S. due to columns = 454.01 – 440.51 = 13.50

Treatment S.S. = 441.22 – 440.51 = 0.71

Error S.S. = Total S.S.–S.S. due to Rows–S.S. due to Columns–Treatment S.S.

= 50.51 – 12.96 – 13.50 – 0.71 = 23.34

(iii) Adjusted treatment S.S.

While the error M.S. is an unbiased estimate of the error variance, the treatment S.S. is an over estimate and has to be corrected by subtracting from it the quantity

$$= \frac{[(t-1)T'+R'+C'-G']^2}{(t-1)^2(t-2)^2} = \frac{[5 \times 16.61+21.03+15.99-122.04]^2}{5^2 \times 4^2}$$

$$= 0.01$$

Therefore, adjusted treatment S.S. = 0.714 – 0.01 = 0.70

(iv) Analysis of variance

ANOVA

Source	d.f.	S.S.	M.S.	F
Rows	5	12.96	2.592	2.111
Columns	5	13.50	2.700	2.198
Treatment (unadjusted)	5	0.71	0.142	<1
Treatment (adjusted)	5	0.70	0.140	<1
Error	19	23.34	1.228	-
Total	34	50.50		

(v) Standard Errors

(a) S.E. of difference between the means not involving missing value

$$SE_d = \sqrt{\frac{2S_e^2}{t}} = \sqrt{\frac{2 \times 1.228}{6}}$$

= 0.640 kg/plot

(b) S.E. of difference between two treatments means one of which involves missing value

$$SE_d = \sqrt{S_e^2\left[\frac{2}{t} + \frac{1}{(t-1)(t-2)}\right]}$$

$$= \sqrt{1.228\left(\frac{2}{6} + \frac{1}{5 \times 4}\right)} = 0.686 \text{ kg/plot}$$

Two missing observation in L.S. Design:

In this case also the method of iteration as illustrated in case of R.B.D. can be used except that X_1 is computed from

$$\hat{X}_1 = \frac{t(R'+C'+T') - 2G'}{(t-1)(t-2)}$$

Standard errors

Case-I

When the observations are missing in different rows, columns and treatments.

(a) S.E. of difference of mean of two affected treatments

$$SE_d = \sqrt{\frac{2(t-2)}{t(t-3)}S_e^2}$$

(b) S.E. of difference of means of an affected treatment and an unaffected one

$$SE_d = \sqrt{\left[\frac{2}{t}+\frac{t(t-3)+2(k-1)}{t[(t-3)[t(t-3)+2k]}\right]S_e^2}$$

Where k, denotes the number of missing observations

Case-II

Where a row, column or treatment is partly, but not completely missing,

(c) S.E. of difference between the means of any two affected treatments when a part of a row or a column is missing but not the part of a treatment.

$$SE_d = \sqrt{\frac{2(t-1)}{t(t-2)}S_e^2}$$

(d) S.E. of difference between the means of an unaffected treatment and that of an affected one.

$$SE_d = \sqrt{\left[\frac{2}{t}+\frac{t-k+1}{t(t-2)(t-k)}\right]S_e^2}$$

(e) S.E. of difference between the means of a treatment which is partly missing and that of an unaffected treatment mean

$$SE_d = \sqrt{\left[\frac{2}{t}+\frac{t}{t(t-2)(t-k)}\right]S_e^2}$$

Standard error of difference for the treatments can also be find

out by

$$SE_d = \sqrt{\left[\frac{1}{r_1} + \frac{1}{r_2}\right] S_e^2}$$

Where r_1 and r_2 are the effective number of replications for the two treatments which are having missing value.

A treatment is given a score 1 if in a row the other treatment is present in row and column, 2/3 if the other treatment is not present in either row or column, 1/3 if the other treatment is not present in both rows and columns and zero if the treatment itself is missing. Let the design be

(A)	(B)	C	D
B	C	D	A
C	D	A	(B)
D	A	(B)	C

For comparing treatments A and B the effective no. of replications for A is: 0+2/3 + 1/3 +1/3 = 4/3

The effective number of replications for B = 0 + 2/3 + 0 + 0 = 2/3

For comparing A and C, the effective no. of replications for A is 0 + 1 +1+1=3, and the effective replication for C is 2/3 + 1 + 2/3 + 1 = 10/3.

For comparing A and D, the effective number of replication for A starting from 1st row = $0 + \frac{2}{3} + \frac{2}{3} + 1 = \frac{7}{3}$

The effective number of replication for

$$D = \frac{2}{3} + 1 + 0 + \frac{2}{3} = \frac{7}{3}$$

$$SE_{(A-D)} = \sqrt{\left(\frac{3}{7} + \frac{3}{7}\right) S_e^2}$$

❐ ❐ ❐

Chapter – 8

Factorial Experiments

In a factorial or a multifactor experiment we study the effect of several factors simultaneously each factor having a number of levels. In these experiments we are interested in estimating and testing the treatment contrasts that define main effects and interaction effects of different factors. Consider the result of a trial conducted to measure the effect of nitrogen and phosphorus on the yield of rice crop. Suppose we have two levels (N_0 and N_1) of nitrogen and two level of phosphorus (P_0 and P_1). The yield of rice crop in q/ha is given as:

Rice Yield in q/ha

Nitrogen **Phosphorus**	**0 kg N/ha (N_0)**		**60 kg N/ha (N_1)**	
0 kg P_2O_5/ha (P_0)	N_0P_0	10.0	N_1P_0	30.0
30 kg P_2O_5/ha (P_1)	N_0P_1	20.0	N_1P_1	40.0

Effect of nitrogen at P_0 level of phosphorus = 30 - 10 = 20.0 q/ha

Effect of nitrogen at P_1 level of phosphorus = 40 - 20= 20 q/ha

Effect of phosphorus at N_0 level of Nitrogen = 20 - 10 = 10 q/ha

Effect of phosphorus at N_1 level of Nitrogen = 40 - 30 = 10 q/ha

As the effect of nitrogen (Phosphorus) is the same at both the levels of phosphorous (nitrogen), there is no interaction between nitrogen and phosphorus. If the yields of rice shows the following pattern.

Rice yield in q/ha

Nitrogen Phosphorus	0 kg N/ha (N_0)		60 kg N/ha (N_1)	
0 kg P_2O_5/ha (P_0)	N_0P_0	10.0	N_1P_0	30.0
30 kg P_2O_5/ha (P_1)	N_0P_1	20.0	N_1P_1	50.0

Effect of nitrogen at P_0 level of phosphorus = 30 - 10 = 20 q/ha

Effect of nitrogen at P_1 level of phosphorus = 50 – 20 = 30 q/ha

Effect of phosphorus at N_0 level of Nitrogen = 20 - 10 =10 q/ha

Effect of phosphorus at N_1 level of Nitrogen = 50 - 30 = 20 q/ha

As the effect of nitrogen (Phosphorus) are not the same at both the level of phosphorus (nitrogen), nitrogen and phosphorus have interaction effect. The effects explained above are called simple effects and average of these simple effects is called main effects of the factor.

For example for the second set of data

Main effect of nitrogen = 20+30/2 = 25.0q/ha.

Main effect of phosphorus = (10.0 +20.0)/2=15.0q/ha

Interaction is the difference between simple effects, thus either 30-20.0=10.0q/ha or 20-10=10q/ha is the measure of interaction effect between nitrogen and phosphorus. It interaction exits which is fairly common, we should plan our experiments in such a way that these can be to estimated and tested. It can not be done if we vary only one factor at a time for this purpose we must have multi-level, multifactor experiments.

The experiment in which several levels (or stages) of two for more than two factors are tested in all possible combinations are called factorial experiments. Here the main effects and interactions are studied together. When the levels of all factors are equal, experiment is called symmetrical factorial experiment otherwise asymmetrical factorial experiment.

In general, 2^n factorial experiments means n factor each at 2 levels, 3^n factorial experiment means n factors each at 3 levels. The

treatment combination, can be laid out in Complete Randomized Design or Randomized Block Design depending upon the nature of the experimental material.

Features of Factorial experiments

(i) It can furnish information regarding the interaction between various factors.

(ii) If the interaction between the factors is significant, the optimum combination of two can be discovered only through factorial combinations.

(iii) If the two factors do not interact, we may speak of the response of one factor irrespective of the levels of the other factors.

Symmetrical Factorial Experiment

2^2 factorial experiment in which two factors A and B each at two levels 0 and 1. There will be four treatment combination.

00 $= a_0\, b_0 = 1$: A and B both are at the first level.

10 $= a_1\, b_0 = a$: A at the second level and B at the first level.

01 $= a_0\, b_1 = b$: A at the first level and B at the second level.

11 $= a_1\, b_1 = ab$: A and B both at the second level.

we denote the treatment combinations by small letters and treatment total by [1], [a], [b], [ab].

The comparison a_1b_1-a_0b_1 estimates the effect of A when B is held constant at higher level. The comparison a_1b_0-a_0b_0 estimates the effect of A when B is held constant at the lower level. The average of these two estimates is called main effect and the difference is called interaction, mathematically.

$$A = \frac{1}{2}(a_1b_1 + a_1b_0 - a_0b_1 - a_0b_0) = \frac{1}{2}(a-1)(b+1)$$

$$B = \frac{1}{2}(a_1b_1 + a_0b_1 - a_1b_0 - a_0b_0) = \frac{1}{2}(a+1)(b-1)$$

$$AB = \frac{1}{2}(a_1b_1 + a_0b_0 - a_1b_0 - a_0b_1) = \frac{1}{2}(a-1)(b-1)$$

How to compute main effects and interaction:

2^2 Factorial Experiment

In this notation, the following table shows how to compute the main effects and interaction from the treatment totals over 'r' replications for 2^2 factorial experiment.

Factorial Effect	Multiplier for Treatment total	Divisor to get mean	S.S.
	(1) a b ab	2r	
A	-1 1 -1 1	2r	$[A]^2/4r$
B	-1 -1 1 1	2r	$[B]^2/4r$
AB	1 -1 -1 1	2r	$[AB]^2/4r$

The rule to write down the signs of the main effects is to give a plus sign to the treatment combination containing the corresponding small letters and a minus sign where the corresponding small letter is absent. The signs of interaction are obtained by multiplying the corresponding signs of the two main effects. The main effect of A is-

$[A]/2r = [(ab) - (b) + (a) - (1)]/2r$

The quantities [A], [B], [AB] are called factorial effect totals.

These four treatments combinations may be compared using either a CRD, RBD or LSD. The analysis will be same as that of the design considered.

Analysis

Let us have 2^2 factorial experiment laid in RBD with 'r' blocks. Thus, the four treatment combination shall be put in 'r' blocks randomly considering the treatment combination as individual treatments. Hence, the ANOVA table will look like as below:

S.E. for testing the difference between means in case of p factor experiment

$$= \sqrt{\frac{2MSE}{r.2^{n-p}}} \quad \forall\ p = 1, 2....n$$

Anova

Source	d.f.	S.S.	M.S.	F
Blocks	r-1	SSB	SSB/r-1	
Main effect A	1	$[A]^2/4r$	MSA	MSA/MSE
Main effect B	1	$[B]^2/4r$	MSB	MSB/MSE
Interaction AxB	1	$[AB]^2/4r$	MSAB	MSAB/MSE
Error	3(r-1)	by subtraction	MSE	
Total	4r-1	$\sum_{i,j} yij^2 - C.F.$		

S.E. main effect = $\sqrt{\frac{MSE}{r}}$, here n=2, p=1

S.E. for interaction AB = $\sqrt{\frac{2MSE}{r}}$ as n=p=2

Critical differences are obtained by multiplying SE by students t value at α level of significance at error degree of freedom.

Yates Method for computing factorial effect totals

Yates gave a method to compute various effect totals for any 2^n-experiment.

Step-1: Write down treatment combinations systematically in the 1^{st} column starting with (1) and then writing a, b, ab and then combinations of these with new symbol c. Again the process is repeated with new symbol d and so on.

Step-2: Write down treatment totals from all the replicates in the second column against appropriate treatment combination.

Step-3: Break the values of the second column in to pairs starting from (1). In column third write down the sums of pairs in the first half while differences in the second half (The first number subtracted from second number)

Step-4: Again break the values of the third column obtained in step 3 into pairs and proceed as in step 3 to get column four and so on.

Remarks: For a 2^n – experiment we are to repeat n- times the operations of column 3 and 4 and then the values in the (n+2)nd column will be factorial effect totals. The first entry will always be grand total.

Yates method for 2^2- experiment

Treat. Comb. (1)	**Total (2)**	**Col. (3)**	**Col. (4)**	**Total effects**	**S.S. due to different effect**
(1)	[1]	[1]+[a]	[1]+[a]+[b]+[ab]	G	$G^2/r.2^n$
a	[a]	[b]+[ab]	[a]-[1]+[ab]-[b]	A	$A^2/r.2^n$
b	[b]	[a]-[1]	[b]+[ab]-[1]-[a]	B	$B^2/r.2^n$
ab	[ab]	[ab]-[b]	[ab]-[b]-[a]+[1]	AB	$(AB)^2/r.2^n$

The analysis of variance can be formed. The procedure is illustrated with an example.

Example: 2^2- experiment in 6 randomized blocks was conducted in order to obtain an idea of the interaction; spacing x number of seedlings per hole, along with the effects of different types of spacing and different numbers of seedlings per hole, while adopting the Japanese method of cultivation of rice. The levels of two factors are:

S : 8" and 10" spacing N : 3 and 4 seedlings per hole.

The field plan and yield is as follows:

B1

(1)	s	ns	n
117	106	109	114

B2

ns	(1)	s	n
114	120	117	114

B3

(1)	n	s	ns
111	117	114	106

B4

ns	n	s	(1)
98	121	112	108

B5

ns	s	(1)	n
75	97	73	38

B6

n	(1)	ns	s
58	81	105	117

Analyse the data to find out if there is any significant treatment effects?

Computations: We apply Yates method to find total effects

Treat. Comb.	Totals yield from all blocks	Column (3)	Column (4)	Total effect	S.S.
(1)	610	1172	2442	GT	$(2442)^2/24=24847.35$
n	562	1270	-104	N	$(-104)^2/24=450.667$
s	663	-48	98	S	$(98)^2/24=400.167$
ns	607	-56	-8	NS	$(-8)^2/24=2.667$

$CF = (2442)^2/24 = 24847.35$

ANOVA can be set as follows:

Source	d.f.	S.S.	M.S.	F
Blocks	5	6270.5	1254.1	1.973
N	1	450.667	450.667	1.752
S	1	400.167	400.167	N.S.
N x S	1	2.667	2.667	N.S.
Error	15	3426.5	228.433	N.S.
Total	23	10550.5		

$F_{(.05;1,15)}$ tabulated = 4.54

Result: There are no significant main or interaction effects present in the above experiment.

2^3 – experiment

If we consider three factors A, B and C each at two levels, then there are 8 treatment combinations denoted by (1), a, b, ab, c, ac, bc, abc. These 8 treatment combinations may be compared in any of the designs CRD, RBD or LSD having the same analysis as if there are 8 treatments.

In a three factor experiment there are three main effects – A, B, C; three first- order interactions – AB, AC, BC; and one second-order (or three factor) interaction-ABC. Here main effect A is defined as

$$A = \frac{1}{4}[(abc)-(bc)+(ac)-(c)+(ab)-(b)+(a)-(1)]$$

$$\text{or } A = \frac{1}{4}(a-1)(b+1)(c+1)$$

$$B = \frac{1}{4}(a+1)(b-1)(c+1)$$

$$AB = \frac{1}{4}(a-1)(b-1)(c+1)$$

Alternatively we can also get the expression for effects by using Table of signs as follows:

Effect	(1)	(a)	(b)	(ab)	(c)	(ac)	(bc)	(abc)	Divisor
M	+	+	+	+	+	+	+	+	8
A	-	+	-	+	-	+	-	+	4
B	-	-	+	+	-	-	+	+	4
C	-	-	-	-	+	+	+	+	4
AB	+	-	-	+	+	-	-	+	4
AC	+	-	+	-	-	+	-	+	4
BC	+	+	-	-	-	-	+	+	4
ABC	-	+	+	-	+	-	-	+	4

Or we can also use Yates method for obtaining the total effects and then proceed for the ANOVA and test the hypothesis.

Example of 2^3 factional experiment

The following table gives the yield of factorial experiment involving 3 factors A, B, C each at two levels laid out in RBD using three replication alongwith total effects obtained by Yates technique. Analyse the data and draw conclusions.

C.F. = $(132)^2/24 = 726.0$

(2) Total sum of square (TSS) = $(2)^2 + (3)^2 \ldots +(14)^2 - 726.0 = 240$

(3) Sum of square due to Blocks = $[(37)^2 +(42)^2 + (53)^2]/8\text{-}776=16.75$

(4) Sum of square due to treatments = 212.68

(5) Sum of square due to error = 240-212.6 – 16.75 = 1057

Table

1	Blocks			2	3	4	5	Total effects	S.S. due to diff. effects
Treatment	B_1	B_2	B_3	Total					
(1)	2	3	3	8	17	44	132	G	726
a	3	2	4	9	27	88	32	(A)	$32^2/24=42.70$
b	2	1	4	7	31	14	36	(B)	$36^2/24=54.0$
ab	5	7	8	20	57	18	24	(AB)	$24^2/24=24.0$
c	5	4	5	14	1	10	44	(C)	$44^2/24=80.70$
ac	4	6	7	17	13	26	4	(AC)	$4^2/24=0.58$
bc	6	7	8	21	3	12	16	(BC)	$16^2/24=10.70$
abc	10	12	14	36	15	12	0	(ABC)	$0^2/24=0.0$
Total	37	42	53						212.68

Anova table

Source	d.f.	S.S.	M.S.	F
Blocks	2	16.75	8.34	12.69**
Treatments				
A	1	42.70	42.70	64.70**
B	1	54.00	54.00	81.82**
AB	1	24.00	24.00	36.36**
C	1	80.70	80.70	122.27**
AC	1	0.58	0.58	0.88
BC	1	10.70	10.70	16.21***
ABC	1	0	0	0
Error	16	10.57	0.66	
Total	23	240.00		

The main effects and the interaction differ significantly except AC and ABC. CD values can be calculated in the usual way as in the RBD.

Disadvantage

When the no. of factors or their levels are large it would result in larger no. of treatment combinations which would require blocks of

larger size. This would result in heterogeneity with in block and the precision of the experiment will be low. Under such situations confounded designs are adopted in which treatments/replicates are divided into two or more blocks of small size and treatments are randomized in each block separately resulting in increasing precision.

In general ANOVA for 2^n experiment is

Source	d.f.
Replication	r-1
Treatment	2^n-1
(i) Main effects	n
(ii) Two factor interaction	$\binom{n}{2}$
(iii) Third factor interaction	$\binom{n}{3}$
(iv) n factor interaction	$\binom{n}{n}$
Error	$(2^n-1)(r-1)$
Total	$r2^n-1$

Asymmetrical factorial experiments

Asymmetrical factorial experiments are those in which all the factors do not have equal no. of levels. These find wide applications in agricultural, industrial and animal science experimentations for testing several factors at unequal no. of levels.

Two factors asymmetrical factorial experiment

Layout: Consider a two factors factorial experiment with 2 varieties and 3 fertilizers laid in R.B.D. with 5 replications. The six possible treatment combinations namely v_1f_1, v_1f_2, v_1f_3, v_2f_1, v_2f_2 and v_2f_3 are randomized in each replicate separately as follows:

R_1	R_2	R_3	R_4	R_5
v_1f_1 117	v_1f_2 171	v_2f_2 61	v_1f_2 146	v_1f_3 137
v_2f_3 74	v_2f_1 48	v_1f_3 125	v_1f_1 112	v_2f_2 54
v_1f_2 139	v_1f_1 143	v_2f_3 62	v_2f_3 76	v_2f_1 22
v_3f_1 35	v_2f_3 97	v_1f_2 138	v_2f_2 55	v_1f_1 93
v_2f_2 44	v_2f_3 155	v_2f_1 36	v_1f_3 147	v_2f_3 48
v_1f_3 129	v_2f_2 71	v_1f_1 133	v_2f_1 42	v_1f_2 122

Record the raw data in the layout plan itself as shown above and rearrange the data in following date sheet for further analysis.

Treatment	**R_1**	**R_2**	**R_3**	**R_4**	**R_5**	**Total**
v_1f_1						v_1f_1
v_1f_2						v_1f_2
v_1f_3						v_1f_3
v_2f_1						v_2f_1
v_2f_2						v_2f_2
v_2f_3						v_2f_3
Total	R_1	R_2	R_3	R_4	R_5	G

Prepare v x f table of total from totals of treatment combinations

	v_1	v_2	Total
f_1	v_1f_1	v_2f_1	F_1
f_2	v_1f_1	v_2f_2	F_2
f_3	v_1f_3	v_2f_3	F_3
Total	V_1	V_2	G

So we have R_1, R_2, ... R_5 are replication totals

V_1, V_2 .....are varieties totals

F_1, F_2, F_3 .....treatments totals

$\Sigma R_i = \Sigma V_j = \Sigma F_k = \Sigma Y = G$ = Grand total

N = rvf = 5 x 2 x 3 = 30 = Total no. of observations

Correction factor (C.F.) = G^2/n

ANOVA Table

Source	d.f.	S.S.	M.S.	F
Replications	r-1 = 4	$\Sigma R_i^2/vf$-C.F.	R	
Varieties	v – 1 = 1	$\Sigma V_i^2/rf$-C.F.	V	V/E~F_{v-1}, error d.f.
Fertilizers	f – 1 = 2	$\Sigma F_k^2/rv$-C.F.	F	F/E~F_{f-1}, error d.f.
Variety x Fertilizer	(v-1) (f-1) = 2	$\Sigma(v_jf_k)2/r$-CF SS(V) – SS(F)	VF	VF/E~$F_{(v-1)(f-1)}$, error d.f.
Error	(fv-1)(r-1)	(by subtraction)	E	
Total	rfv-1	ΣY^2 - CF		

Using F test significance of each component namely variety, fertilizer and their interaction variety x fertilizer can be tested against error as shown above,

SE of difference for varieties = $\sqrt{2E/rf}$

SE of difference fertilizers = $\sqrt{2E/rv}$

SE of difference for V x F = $\sqrt{2E/r}$

Now C.D. value can be calculated by multiplying the SE_d with $t_{(\alpha,\ \text{error d.f.})}$ value

For testing the levels for different factors and interaction, prepare V x F table of means by dividing each events of V x F table of total by r and compare the various factors using respective 5% C.D.

V x F mean table

	v_1	v_2	**Totals**
f_1	v_1f_1/r	v_2f_1/r	f_1
f_2	-	-	f_2
f_3	-	v_2f_3/r	f_3
Totals	v_1	v_2	

Example: Give complete analysis of factorial experiment (given in layout) on 2 varieties, 3 fertilizers conducted in RBD with 5 replications.

Solution:

Arrange the data in following datasheet and find marginal totals.

Replications

Treatment	I	II	III	IV	V	Total
v_1f_1	117	143	133	112	93	598
v_1f_2	139	171	138	146	122	716
v_1f_3	129	155	125	147	137	693
v_2f_1	35	48	36	42	22	183
v_2f_2	44	71	61	55	54	285
v_2f_3	74	97	62	76	48	357
Total	538	685	555	578	476	2832

Now prepare V x F table of totals from above treatment totals and find totals on both sides

	v_1	v_2	Totals
f_1	598	183	781
f_2	716	285	1001
f_3	693	357	1050
Totals	2007	825	2832

We have r =5, v=2, f=3, t=vf=6, N=rvf=30 G=2832

$$C.F. = G^2/N = \frac{(2832)^2}{30} = 267340.8$$

$$\text{SS due to replicates} = \frac{\Sigma R_i^2}{vf} - C.F.$$

$$= \frac{(538)^2 + (685)^2 + ...(476)^2}{6} - 267340.8 = 2689.1$$

$$\text{SS due to varieties} = \frac{\Sigma Vj^2}{rf} - C.F.$$

$$= \frac{(2007)^2 + (825)^2}{15} - 267340.8 = 46570.8$$

$$\text{SS due to fertilizers} = \frac{\Sigma Fk^2}{rv} - \text{C.F.}$$

$$= \frac{(781)^2 + (1001)^2 + (1050)^2}{10} - 267340.8 = 4105.4$$

$$\text{S.S. due to interaction V x F} = \frac{\Sigma(vjfk)^2}{r} - \text{C.F.} - SS(V) - SS(F)$$

$$= \frac{(598)^2 + (716)^2 + \ldots + (357)^2}{5} - 267340.8$$

$$-46570.8 - 4105.4 = 517.4$$

$$\text{Total SS} = \Sigma Y^2\text{-C.F.} = (117)^2 + (139)^2 + \ldots\ldots(48)^2 - 267340.8$$

$$= 56675.2$$

ANOVA Table

Source	df	S.S.	M.S.	F
Replicates	r-1=4	2689.1	672.28	4.8
Varieties	v-1=1	46570.8	46570.80	333.5*
Fertilizer	f-1=2	4105.4	2052.73	14.7*
Varity x Fertilizer	(v-1) (f-1) = 2	517.4	258.72	1.9
Error	20	2792.5	139.63	
Total	rvf-1 = 29	56675.2		

$F_{(.05,\ 1,\ 20)} = 4.35,$ $F_{(.05,\ 2,\ 20)} = 3.49$

Since for both varieties and fertilizer, F_{cal} > F tabulated, so they differ significantly, while their interaction is non significant.

$$\text{C.D. for varieties at 5\%} = \sqrt{\frac{2E}{rf}} \times t_0 = \sqrt{\frac{2 \times 139.63}{15}} \times 2.09$$

$$= 4.13 \times 2.09 = 9.01$$

$$\text{C.D. for Fertilizer at 5\%} = \sqrt{\frac{2E}{rv}} \times t_0 = \sqrt{\frac{2 \times 139.63}{10}} \times 2.09$$

$= 5.28 \times 2.09 = 11.04$

Now from VxF table of totals prepare VxF mean table by dividing each entry in table by 5 and taking means on both sides.

	v_1	v_2	Means
f1	119.6	36.6	78.1
f2	143.2	57.0	100.1
f3	138.6	71.4	105.0
Means	133.8	55.0	

CD at 5% for V = 9.01, CD at 5% for F = 11.04 CD for VF = N.S.

Remarks

I. C.D. should not be worked out for the components indicated non significant by F-test in ANOVA table and their means should not be compared otherwise misleading results are obtained.

II. Take 1/F instead of F value when F_{cal} for a component in less than 1. Also interchange the d.f. of F for testing the significance.

Three factor asymmetrical factorial experiment:

It is an extension of two factor factorial experiments.

Layout: Consider a three factor factorial experiment with 2 varieties (v_1,v_2), 3 fertilizers (f_1, f_2, f_3) and two pesticides (p_1, p_2) laid in RBD with r = 4 replications.

Total number of treatment combination t = vfp = 2 x 3 x 2 = 12

Total number of observation N = rvfp = 48

The experiment is conducted in RBD with following 12 treatment combinations.

$t_1=v_1f_1p_1$	$t_7=v_2f_1p_1$
$t_2=v_1f_1p_2$	$t_8=v_2f_1p_2$
$t_3=v_1f_2p_1$	$t_9=v_2f_2p_1$
$t_4=v_1f_2p_2$	$t_{10}=v_2f_2p_2$
$t_5=v_1f_3p_1$	$t_{11}=v_1f_3p_1$
$t_6=v_1f_3p_2$	$t_{12}=v_2f_3p_2$

Analysis

The raw data from the layout plan for further analysis be arranged in the following data sheet.

Treatments	R_1	R_2	R_3	R_4	Total
$v_1f_1p_1$					$v_1f_1p_1$
$v_1f_1p_2$					$v_1f_1p_2$
$v_1f_2p_1$					.
$v_1f_2p_2$					.
$v_1f_3p_1$					.
$v_1f_3p_2$					.
$v_2f_1p_1$					.
$v_2f_1p_2$					.
$v_2f_2p_1$					.
$v_2f_2p_2$					.
$v_1f_3p_1$					.
$v_2f_3p_2$					$v_2f_3p_2$
Total	R_1	R_2	R_3	R_4	

From the total of various treatment combinations construct PxV, VxF and FxP tables of totals

P x V				V x F					F x P			
	v_1	v_2	Total		f_1	f_2	f_3	Total		p_1	p_2	Total
p_1	v_1p_1	v_2p_1	P_1	v_1	v_1f_1	v_1f_2	v_1f_3	V_1	f_1	f_1p_1	f_1p_2	F_1
p_2	v_1p_2	v_2p_2	P_2	v_2	v_2f_1	v_2f_2	v_2f_3	V_2	f_2	f_2p_1	f_2p_2	F_2
Total	V_1	V_2			F_1	F_2	F_3		f_3	f_3p_1	f_3p_2	F_3
									Total	P_1	P_2	

Hence we have R_1, R_2, R_3, R_4 as replication totals

V_1, V_2 are varieties totals

F_1, F_2, F_3, are fertilizer totals

P_1, P_2 are pesticide totals

$\Sigma R_i = \Sigma V_j = \Sigma F_k = \Sigma P_e = G$ = Grand total

N = r v f p = Total number of observation

Correction factor (C.F.) = G^2/N

ANOVA Table

Source	d.f.	S.S.	M.S.	F
Replications	r-1=3	$\frac{\Sigma R_i^2}{vfp} - CF$	R	
Varieties	v-1=1	$\frac{\Sigma V_j^2}{rfp} - CF$	V	V/E
Fertilizer	f-1=2	$\frac{\Sigma F_k^2}{rvp} - CF$	F	F/E
Variety x Fertilizer	(v-1) (f-1) =2	$\frac{\Sigma\Sigma(v_j f_k)^2}{rp} - CF - SS(V) - SS(F)$	VF	VF/E
Pesticides	p-1=1	$\Sigma P_l^2 / rvf - CF$	P	P/E
Variety x pesticides	(v-1) (p-1)=1	$\frac{\Sigma\Sigma(v_j p_l)^2}{rf} - CF - SS(V) - SS(P)$	VP	VP/E
Fertilizer x pesticides	(f-1) (p-1)=2	$\frac{\Sigma\Sigma(f_k p_l)^2}{rv} - CF - SS(F) - SS(P)$	FP	FP/E
V x F x P	(v-1) (f-1) (p-1) = 2	$\frac{\Sigma\Sigma(v_j f_k p_l)^2}{r} - CF - SS(V) - SS(F)$ -SS(VxF)-SS(P)-SS(VxP)-SS(FxP)	VFP	VFP/E
Error	By subtraction	By subtration	E	
Total	rvfp-1	Σy^2-CF		

All the main effects and interactions are tested against error E using F-test with their respective d.f.

5% CD for any main effect or interaction is given by $\sqrt{2E/r_i^*} \times t_0$

Where r_i^* = effective no. of replications for that component and is equal to the divisor of respective component of S.S. in the ANOVA table.

For further comparison prepare PxV, VxF and FxP tables of mean by dividing the entries in table of totals by r_i^* where ri* is the

effective number of replications of that interaction.

P x V				V x F				F x P		
	v_1	v_2			f_1	f_2	f_3		p_1	p_2
p_1				v_1				f_1		
p_1				v_2				f_2		
Means	v_1	v_2		Means	f_1	f_2	f_3	f_3		
								Means	p_1	p_2

The significance of difference between varieties, fertilizers, pesticides and their interaction can be tested from above mean tables using the corresponding C.D. values.

Experiments with factors each at three levels

When factors are taken at three levels instead of two, the scope of an experiment increases. It becomes more informative. A study to investigate if the change is linear or quadratic is possible when the factors are at three levels. The more the number of levels, the better, yet the number of the levels of the factors cannot be increased too much as the size of the experiment increases too rapidly with them. Let us begin with two factors A and B, each at three levels say 0, 1 and 2(3^2 –factorial experiment). The treatment combinations are

00 $= a_0b_0 = 1$; A and B both at first levels

10 $= a_1b_0 = a$; A is at second level and B is at first level

20 $= a_2b_0 = a^2$; A is at third level and b is at first level

01 $= a_0b_1 = b$; A is at first level and B is at second level

11 $= a_1b_1 = ab$; A and B both at second level

21 $= a_2b_1 = a^2b$; A is at third level and B is at second level

02 $= a_0b_2 = b^2$; A is at first level and B is at third level

12 $= a_1b_2 = ab^2$; A is at second level and B is at third level

22 $= a_2b_2 = a^2b^2$; A and B both at third level

Any standard design can be adopted for the experiment.

The main effects A, B can respectively be divided into linear and

quadratic components each with 1 d.f. as A_L, A_Q, B_L and B_Q. Accordingly AB can be partitioned into four components as A_LB_L, A_LB_Q, A_QB_L, A_QB_Q.

The coefficients of the treatment combinations to obtain the above effects are given as

Treatment totals Factorial Effects	[1]	[a]	[a^2]	[b]	[ab]	[a^2b]	[b^2]	[ab^2]	[a^2b^2]	Divisor
M	+ 1	+1	+1	+1	+ 1	+1	+1	+1	+ 1	$9r = rx3^2$
A_L	-1	0	+1	- 1	0	+1	-1	0	+ 1	6r = rx2x3
A_Q	+1	-2	+1	+1	-2	+1	+1	-2	+1	18r = 6x3
B_L	-1	-1	-1	0	0	0	+1	+1	+1	6r = rx2x3
A_LB_L	+1	0	-1	0	0	0	-1	0	+1	4r = rx2x2
A_QB_L	-1	+2	-1	0	0	0	+1	-2	+1	12r = rx6x2
B_Q	+1	+1	+1	-2	-2	-2	+1	+1	+1	18r = rx3x6
A_LB_Q	-1	0	+1	+2	0	-2	-1	0	+1	12r = rx2x6
A_QB_Q	+1	-2	+1	-2	+4	-2	+1	-2	+1	36r = rx6x6

The rule to write down the coefficients of the linear (quadratic) main effects is to give a coefficient as +1 (+1) to those treatment combinations containing the third level of the corresponding factor, coefficient as 0 (-2) to the treatment combinations containing the second level of the corresponding factor and coefficient at -1(+1) to those treatment combinations containing the first level of the corresponding factor. The coefficients of the treatment combinations for two factor interactions are obtained by multiplying the corresponding coefficients of two main effects. The various factorial effect totals are given as

$[A_L] = +1[a^2b^2] + 0[ab^2] - 1[b^2] + 1[a^2b] + 0[ab] - 1[b] + 1[a^2] + 0[a] - 1[1]$

$[A_Q] = +1[a^2b^2] - 2[ab^2] + 1[b^2] + 1[a^2b] - 2[ab] + 1[b] + 1[a^2] - 2[a] + 1[1]$

$[B_L] = +1[a^2b^2] + 1[ab^2] + 1[b^2] + 0[a^2b] + 0[ab] + 0[b] - 1[a^2] - 1[a] - 1[1]$

$[A_LB_L] = +1[a^2b^2] + 0[ab^2] - 1[b^2] + 0[a^2b] + 0[ab] + 0[b] - 1[a^2] + 0[a] - 1[1]$

$[A_QB_L] = +1[a^2b^2] - 2[ab^2] + 1[b^2] + 0[a^2b] + 0[ab] + 0[b] - 1[a^2] + 2[a] - 1[1]$

$[B_Q] = +1[a^2b^2] + 1[ab^2] + 1[b^2] - 2[a^2b] - 2[ab] - 2[b] - 1[a^2] - 1[a] - 1[1]$

$[A_LB_Q] = +1[a^2b^2] + 0[ab^2] - 1[b^2] - 2[a^2b] + 0[ab] + 2[b] + 1[a^2] + 0[a] - 1[1]$

$[A_QB_Q] = +1[a^2b^2] - 2[ab^2] + 1[b^2] - 2[a^2b] + 4[ab] - 2[b] + 1[a^2] - 2[a] + 1[1]$

Factorial effects are given by

$A_L = [A_L]/r.3$ $A_Q = [A_Q]/r.3$ $B_L = [B_L]/r.3$ $A_LB_L = [A_LB_L]/r.3$

$A_QB_L = [A_QB_L]/r.3$ $B_Q = [B_Q]/r.3$ $A_LB_Q = [A_LB_Q]/r.3$ $A_QB_Q = [A_QB_Q]/r.3$

The sum of squares due to various factorial effects is given by

$$SSA_L = \frac{[A_L]^2}{r.2.3} \quad SSA_Q = \frac{[A_Q]^2}{r.6.3} \quad SSB_L = \frac{[B_L]^2}{r.3.2} \quad SSA_LB_L = \frac{[A_LB_L]^2}{r.2.2}$$

$$SSA_QB_L = \frac{[A_QB_L]^2}{r.6.2} \quad SSB_Q = \frac{[B_Q]^2}{r.3.6} \quad SSA_LB_Q = \frac{[A_LB_Q]^2}{r.2.6} \quad SSA_QB_Q = \frac{[A_QB_Q]^2}{r.6.6}$$

If a randomized complete block design is used with r-replications then the outline of analysis of variance is

Anova

Source of variation	**degree of freedom**	**S.S.**	**M.S.**
Between replications	r-1	SSR	MSR=SSR/(r-1)
Between treatments	$3^2 - 1 = 8$	SST	MST=SST/8
A	2	SSA	MSA=SSA/2
A_L	1	SSA_L	$MSA_L = SSA_L$
A_Q	1	SSA_Q	$MSA_Q=SSA_Q$
B	2	SSB	MSB=SSB/2
B_L	1	SSB_L	$MSB_L = SSB_L$
B_Q	1	SSB_Q	$MSB_Q=SSB_Q$
AB	4	SSAB	MSAB=SSAB/2
A_LB_L	1	SSA_LB_L	$MSA_LB_L=SSA_LB_L$
A_QB_L	1	SSA_QB_L	$MSA_QB_L=SSA_QB_L$
A_LB_Q	1	SSA_LB_Q	$MSA_LB_a=SSA_LB_Q$
A_QB_Q	1	SSA_QB_Q	$MSA_aB_a=SSA_QB_Q$
Error	$(r-1)(3^2-1) = 8(r-1)$	SSE	MSE=SSE/8(r-1)
Total	$r.3^2-1=9r-1$	TSS	

If the factors are quantitative in nature, retain the SS with single d.f.. However, if the factors are qualitative in nature, then this partitioning of SS of a particular factorial effect with p d.f. can only be

considered as partition of its SS into single d.f. components by writing 'p' linearly independent contrasts and these sum of, squares with single d.f. can be added to obtain the SS due to corresponding factorial effect totals with p d.f.

In general, for n factors each at 3 levels, the sum of squares due to any linear (quadratic) main effect is obtained by dividing the square of the linear (quadratic) main effect total by $r.2.3^{n-1}$($r.6.3^{n-1}$). Sum of squares due to a 'p' factor interaction is given by taking the square of the total of the particular interaction component divided by $r.(a_1\ a_2 \ldots a_p).\ 3^{n-p}$, where $a_1, a_2, \ldots\ldots, a_p$ are taken as 2 or 6 depending upon the linear or quadratic effect of particular factor.

Exercise 2: A 3^2 experiment was conducted to study the effects of the factors Nitrogen (N) and Phosphorus (P) (each at three levels 0,1,2) on sugar beet. Two replications of nine plots each were used. The table below shows the plan and the percentage of sugar (approximated to nearest whole number).

Replication	Treatment		% of sugar
	N	P	
I	0	1	14
	2	0	15
	0	0	16
	2	1	15
	0	2	16
	1	2	18
	1	1	17
	1	0	19
	2	2	17
II	1	2	20
	1	0	19
	1	1	17
	0	0	18
	2	1	19
	0	1	16
	0	2	16
	2	2	19
	2	0	16

Analysis

Step 1: To find sum of squares between replications, between treatments and total sum of squares arrange the data in a Replication x Treatment combinations table as follows:

Repl.	Treatment Combinations									Total
	1	n	n^2	p	np	n^2p	p^2	np^2	n^2p^2	
	00	10	20	01	11	21	02	12	22	
1	16	19	15	14	17	15	16	18	17	147 (R_1)
2	18	19	16	16	17	19	16	20	19	160 (R_2)
Total	34	38	31	30	34	34	32	38	36	307
	(T_1)	(T_2)	(T_3)	(T_4)	(T_5)	(T_6)	(T_7)	(T_8)	(T_9)	(G)

Grand Total : 307

No. of observations (n) = $r.3^2$ =18

Correction Factor (C.F.)= $(307)^2/18$ =5236.0556

Total S.S. (TSS) = Sum $(\text{observation})^2$ – C.F.

$= 16^2 + 18^2 + \ldots\ldots + 17^2 + 19^2 - 5236.0556$

= 48.9444

$$\text{Replication SS (SSR)} = \frac{R_1^2 + R_2^2}{9} - \text{C.F.}$$

$$= \frac{147^2 + 160^2}{9} - 5236.0556 = 9.3888$$

$$\text{Treatment SS (SST)} = \frac{\text{Sum}(\text{treatment totals})^2}{r} - \text{C.F.}$$

$$= \frac{34^2 + 38^2 \ldots\ldots + 38^2 + 36^2}{2} - 5236.0556 = 32.4444$$

Error SS = Total SS – Replication SS – Treatment SS

= 7.1112

Step2: Obtain various factorial effects total

$[N_L] = +1[n^2p^2] + 0[np^2] - 1[p^2] + 1[n^2p] + 0[np] - 1[p] + 1[n^2] + 0[n] - 1[1]$

$= 5$

$[N_Q] = +1[n^2p^2] - 2[np^2] + 1[p^2] + 1[n^2p] - 2[np] + 1[p] + 1[n^2] - 2[n] + 1[1]$

$= -23$

$[P_L] = +1[n^2p^2] + 1[np^2] + 1[p^2] + 0[n^2p] + 0[np] + 0[p] - 1[n^2] - 1[n] - 1[1] = 3$

$[N_LP_L] = +1[n^2p^2] + 0[np^2] - 1[p^2] + 0[n^2p] + 0[np] + 0[p] - 1[n^2] + 0[n] - 1[1]$

$= 7$

$[N_QP_L] = +1[n^2p^2] - 2[np^2] + 1[p^2] + 0[n^2p] + 0[np] + 0[p] - 1[n^2] + 2[n] - 1[1]$

$= 3$

$[P_Q] = +1[n^2p^2] + 1[np^2] + 1[p^2] - 2[n^2p] - 2[np] - 2[p] - 1[n^2] - 1[n] - 1[1] = 13$

$[N_LP_Q] = +1[n^2p^2] + 0[np^2] - 1[p^2] - 2[n^2p] + 0[np] + 2[p] + 1[n^2] + 0[n] - 1[1]$

$= -7$

$[N_QP_Q] = +1[n^2p^2] - 2[np^2] + 1[p^2] - 2[n^2p] + 4[np] - 2[p] + 1[n^2] - 2[n] + 1[1]$

$= -11$

Step 3: The sum of squares due to various factorial effects is given by

$$SSN_L = \frac{[N_L]^2}{r.2.3} = \frac{[5]^2}{12} = 2.0833$$

$$SSN_Q = \frac{[N_Q]^2}{r.6.3} = \frac{[-23]^2}{36} = 14.6944$$

$$SSP_L = \frac{[P_L]^2}{r.3.2} = \frac{[3]^2}{12} = 0.7500$$

$$SSN_LP_L = \frac{[N_LP_L]^2}{r.2.2} = \frac{[7]^2}{8} = 6.1250$$

$$SSN_QP_L = \frac{[N_QP_L]^2}{r.6.2} = \frac{[3]^2}{24} = 0.375$$

$$SSP_Q = \frac{[P_Q]^2}{r.3.6} = \frac{[13]^2}{36} = 4.6944$$

$$SSN_L P_Q = \frac{[N_L P_Q]^2}{r.2.6} = \frac{[-7]^2}{24} = 2.0417$$

$$SSN_Q P_Q = \frac{[N_Q P_Q]^2}{r.6.6} = \frac{[-11]^2}{72} = 1.6806$$

Step 4: Construct the ANOVA table as given above and test the significance of the various factorial effects.

Anova

Source of variation	degree of freedom	S.S.	M.S.	F tabulated
Between replications	1	9.3888	9.3888	10.5623*
Between treatments	8	32.4444	4.0555	4.5624*
N	2	16.7774	8.3887	9.4371*
N_L	1	2.0833	2.0833	2.3437
N_Q	1	14.6944	14.6944	16.5310*
P	2	5.4444	2.7222	3.0624
P_L	1	0.7500	0.7500	0.8437
P_Q	1	4.6944	4.6944	5.2811
NP	4	10.2223	2.5556	2.875
$N_L P_L$	1	6.1250	6.1250	6.8905*
$N_Q P_L$	1	0.3750	0.3750	0.4219
$N_L P_Q$	1	2.0417	2.0417	2.2968
$N_Q P_Q$	1	1.6806	1.6806	1.8906
Error	8	7.1112	0.8889	
Total	17	48.9444		

(* indicates the significance at 5% level of significance)

❐ ❐ ❐

Chapter – 9

Confounding

When in a factorial experiment, the number of treatment or the levels of the treatment increases, the size of the block increases, resulting in heterogeneity of soil within the block which contributes to the increase in experimental error. For example, 2^5 factorial experiments would have 32 treatment combinations and blocks of 32 plots are quite big to ensure homogeneity within them. A new technique is therefore necessary for designing experiments with a large number of treatments and one such device is to take blocks of size less than the number of treatments and have more than one block per replication. The division of complete block in to incomplete blocks is done in such a way that certain interaction effect (preferably higher order interactions) would be made identical to the incomplete block comparison. In such cases interaction effect is known to be confounded with block effects. This device of reducing the block size by taking one or more interaction contrasts identical with block contrasts is known as confounding. Preferably, only higher order interaction, that is, interactions with three or more factors are to be confounded because their loss is immaterial, as an experimenter is generally interested in main effects and two factor interactions, these should not be confounded as far as possible.

When there are two or more replications, if the same set of interactions are confounded in all the replications, confounding is said to be complete and if a different set of interactions are

confounded in different replications, confounding is said to be partial. In complete confounding all the information on the confounded interactions are lost. But in partial confounding information on the confounded interactions can be recovered from those replications in which they are not confounded.

For example consider 2^3 factorial experiment let the three factors are A, B and C each at two levels.

Let us divide the 8 treatments combinations in two blocks of size 4.

Block I
(1)
(ab)
(ac)
(bc)

Block II
(a)
(b)
(c)
(abc)

Here the Block effect $(B_2 - B_1)$= (abc)+(a)+(b)+(c)–(1)–(ab)–(ac)-(bc) is identical with the interaction ABC = (a-1) (b-1) (c-1).

Here we say ABC effect is confounded or mixed up with the block effect and can not be isolated from the block effect.

Advantages of confounding

If reduces the experimental error considerably by stratifying the experimental material in to homogeneous groups.

Disadvantage of confounding

(1) Here the increased precision is obtained at the cost on certain relativity unimportant interactions.

(2) Confounded contrasts are replicated fewer times than are the other contrasts and as such there is loss of information on them and they can be estimated with a lower degree of precision as the number of replication for them is reduced.

Complete Confounding:

Let us have three factors A, B, C each at two levels. ABC interaction is confounded in all the replications. Layout of 2^3

experiments in blocks of 4 units where ABC is completely confounded in all the three replications.

Rep	R-I		R-II		R-III	
Blocks	B_1	B_2	B_3	B_4	B_5	B_6
	ab	b	ac	c	(1)	abc
	ac	c	ab	b	ab	a
	bc	abc	(1)	a	bc	c
	(1)	a	bc	abc	ac	b

For all the three replicates $B_2 - B_1 = B_4 - B_3 = B_6 - B_5 = ABC$ interaction. Hence ABC is confounded in all the replicates as ABC effect is equal to Block effect in each replication.

Analysis

For carrying out the statistical analysis, the various factorial effects and their S.S. are estimated in the usual manner using Yates technique of the factorial experiments with the modification that for completely confounded interactions neither the S.S. due to confounded interaction is computed nor it is included in the ANOVA table. The confounded component is contained in the (2p-1) d.f. (in case of p replicates) due to blocks. The partitioning of the d.f. for 2^3 completely confounded factorial is as follows:

Source of Variation	Degrees of Freedom
Blocks	2p-1
A	1
B	1
C	1
AB	1
AC	1
BC	1
Error	6(p-1)
Total	8p-1

In general for a $(2^n, 2^k)$ completely confounded factorial in p replications, the different d.f.'s are given as follows:

Source of Variation	Degrees of Freedom
Replication	p-1
Blocks within Replication	$p(2^{n-k}-1)$
Treatment	$2^n-1-(2^{n-k}-1)$
Error	By subtraction
Total	$p2^n-1$

The treatment d.f. has been reduced by $2^{n-k}-1$ as this is the total d.f. confounded per replication.

Partial Confounding

Here we confound different effects in different replications, so that we can get information on each effect from the replications in which the effect is not confounded. Here the effects are said to be partially confounded.

Example

R-I

B_1	B_2
ab	b
ac	c
bc	abc
(1)	a

ABC is confounded

R-II

B_3	B_4
ac	c
b	ab
abc	bc
(1)	a

AC is confounded

R-III

B_5	B_6
ab	b
c	ac
abc	bc
(1)	a

AB is confounded

General method of construction of confounded Designs in 2^n factorial experiments:

Let the design is $(2^n, 2^k)$ i.e. 2^n treatment combinations arranged in 2^k plots per block.

Treatment combinations = 2^n.

Block size = 2^k

Number of blocks per replication = 2^{n-k}

Total number of interaction confounded = $2^{n-k}-1$

Number of independent interaction confounded = (n-k)

Generalized interaction confounded = $(2^{n-k}-1)-(n-k)$

(i) Choose (n-k) independent effect to be confounded. In general these should be of higher interaction of lesser importance. Their generalized interaction should not contain main effects or two factor interactions.

(ii) Choose k independent treatment combinations having even number of letters common with the (n-k) independent confounded effects. (0 is also taken as even number). Denote them by 1, 2, ..., k.

(iii) To find out the other contents of the blocks take 1 x 2, 1 x 3, 2 x 3, 1 x 2 x 3 and delete a^2 b^2 etc. when so ever appears. The process will provide 2^k-1 treatment combination of the block.

(iv) Add control (1) to complete the 2^k treatment combination of key block. This block will be known as key block.

(v) Multiply the contents of the key block by a treatment combination which has not appeared in the key block. As done earlier delete a^2b^2 etc. when so ever appear. This process will provide 2^k treatment combinations of the IInd block.

(vi) Like wise multiply the contents of the first block by a treatment combination which has not appeared in Ist or IInd block earlier to obtain the contents of the third block.

(vii) The process is repeated till all the desired blocks are obtained.

Example

Consider (2^4, 2^2) factorial experiment in two replications. Let us confound (n-k) independent interaction in each replicate. So 2 independent interactions are confounded.

Let in rep I ABC and AD are confounded. So the generalized ABC x AD= A^2BCD = BCD is also confounded.

Let in rep II BC and ABD are confounded. The Generalized interaction BC x ABD = AB^2CD = ACD is also confounded.

No. of plots per replicate in = $2^{4-2} = 2^2 = 4$.

Find k i.e. 2 independent treat combination which have even no.

of letter common with the confounded interaction ABC and AD, they can be bc, and abd and obtained the key block as given in the general method of construction of design.

Rep-I

(1)	bc	bcd	(1)	abc
(2)	abd	ab	abcd	bd
(3) = (1 x 2)	acd	ac	ad	cd
(4)	(1)	d	c	a
	B_1	B_2	B_3	B_4

The other block can be obtained from the key block by multiplying the content of the key block by d, c and a respectively.

Similarly the Block content of the Rep-II can be obtained. Find out two independent treatment combination let ad and abc which have even number of letters common with the confound interactions BC and ABD. The key block and rest of the block can be formed in the same way as given in the general method of construction.

(1)	ad	abd	acd	a
(2)	abc	ac	ab	abcd
(3) = (1 x 2)	bcd	cd	bd	bc
(4)	(1)	(b)	c	d
	B_1	B_2	B_3	B_4

Partial confounding

In case of partial confounding, the S.S. for all the effects, other than the confound one, are obtained as usual using Yates technique given in the factorial experiments. For confounded effects, since the effect is not confounded in all the replications, therefore we can estimate the effects confounded in one replication from the other replication in which it is not confounded. In $(2^n,2^k)$ factorial experiment with p replications, following is the splitting of d.f.'s.

Source of Variation	Degrees of Freedom
Replication	p-1
Blocks within Replication	$p(2^{n-k}-1)$
Treatment	2^n-1
Error	By subtraction
Total	$p2^n-1$

The S.S. for confounded effects are to be obtained from those replications only in which the given effect is not confounded i.e. in calculating the S.S., the observation pertaining to the replication in which it is confounded is not considered.

The S.S. due to confounded interaction is given by

$S.S. = [Z]^2/(r-k^*)2^n$

where Z is the interaction confounded and k* is the number of replication in which the given interaction is confounded.

Example

Analyse the following 2^3 Factorial-experiment in blocks of 4 plots, involving three fertilizers N, P, K each at two levels.

Replication I		Replication II		Replication III	
Block 1	**Block 2**	**Block 3**	**Block 4**	**Block 5**	**Block 6**
np	p	(1)	np	pk	n
101	88	125	115	75	53
npk	n	npk	k	nk	npk
111	90	95	95	100	76
(1)	pk	nk	pk	(1)	p
75	115	80	90	55	65
k	nk	p	n	np	k
55	75	100	80	92	82

Step 1: Identify the interactions confounded in each replicate. Here, each replicate has been divided into two blocks, one effect has been confounded in each replicate. The effects confounded are

Replicate I : NP Replication II : NK Replicate III : NPK

Step 2: Obtain the C.F., blocks S.S. and Total S.S.

$$C.F. = (2088)^2/24 = 181656$$

$$\text{S.S. due to Blocks} = \sum_{i=1}^{6} \frac{B_i^2}{4} - C.F. = 2506$$

$$\text{Total S.S.} = \Sigma(\text{Obs.})^2 - C.F. = 8658$$

Step 3: Obtain the sum of squares due to all the factorial effects other than the confounded effects using Yates technique.

Treatment Combinations	Total Yield	Factorial Effects	Sum of Squares (S.S.) = [ ]$^2/2^3.r$
(1)	255	G=2088	
n	223	[N]=48	$96 = S_N^2$
p	253	[P]=158	$1040.17 = S_P^2$
np	308	[NP]=66	-
k	232	[K]=10	$4.17 = S_k^2$
nk	255	[NK]=2	-
pk	280	[PK]=-8	$2.67 = S_{Pk}^2$
npk	282	[NPK]=-108	-

Step 3: Obtain the S.S. of the confounded interactions. The interaction NP is given by

$$[NP] = [npk] + [np] + [k] + [1] - [pk] - [nk] - [p] - [n] = 92$$

This total is obtained by excluding the observations in the first replicate as NP is confounded in this replicate. Similarly [NK] is obtained from replications I and III and [NPK] is obtained from replications I and II.

$$[NK] = [npk] + [p] + [nk] + [1] - [pk] - [np] - [k] - [n] = -18$$

$$[NPK] = [npk] + [p] + [k] + [n] - [np] - [nk] - [pk] - [1] = -62$$

The S.S. are obtained as follows.

$$S_{NP}^2 = \text{S.S. due to NP} = \frac{1}{16}[NP]^2 = 529;$$

$$S^2_{NK} = \text{S.S. due to NK} = \frac{1}{16}[NK]^2 = 20.25;$$

$$S^2_{NPK} = \text{S.S. due to NPK} = \frac{1}{16}[NPK]^2 = 240.25$$

$$\text{Treatment S.S.} = S^2_N + S^2_P + S^2_K + S^2_{NP} + S^2_{NK} + S^2_{PK} + S^2_{NPK}$$

$$= 1932.7501$$

ANOVA

Source of Variation	d.f.	Sum of Squares	M.S.	Variance Ratio 'F'
Block	5	2506	501	1.31
Treatments	7	1932.75	276.107	-
N	1	96.00	96.00	-
P	1	1040.16	1040.16	2.71
NP	1	529.00	529.00	1.3
K	1	4.41	4.41	-
NK	1	20.25	20.25	-
PK	1	2.66	2.66	-
NPK	1	240.25	240.25	-
Error	11	4219.24	383.57	-
Total	23	8658		

'-' indicates that these ratios are less than one and hence these effects are non-significant.

From the above table it is seen that effects due to blocks, main effects due to factor N, P, and K and interactions are not significant.

Given a set of interactions confounded, how to obtain the blocks?

The blocks of the design pertaining to the confounded interaction can be obtained by solving the equations obtained from confounded interaction.

Example:

Construct a 2^5 factorial in 2^3 plots per block confounding interactions ABD, ACE and BCDE.

Let x_1, x_2, x_3, x_4 and x_5 denote the levels (0 or 1) of each of the 5

factors A,B,C,D and E. Solving the following equations (mod 2) would result in different blocks of the design.

For interaction ABD: $x_1 + x_2 + x_4 = 0, 1$

For interaction ACE: $x_1 + x_3 + x_5 = 0, 1$

BCDE is the interaction that gets confounded automatically and it is called as the generalized interaction. Treatment combinations satisfying the following solutions of above equations will generate the required four blocks

(0, 0) (0, 1) (1, 0) (1, 1)

The solution (0, 0) will give the key block (A key block is one that contains one of the treatment combination of factors, each at lower level).

There will be $\frac{2^5}{2^3}$ =4 blocks per replicate. The key block is as obtained below:

A	B	C	D	E	
1	1	1	0	0	abc
1	1	0	0	1	abe
1	0	1	1	0	acd
1	0	0	1	1	ade
0	1	1	1	1	bcde
-0	1	0	1	0	bd
0	0	1	0	1	ce
0	0	0	0	0	(1)

Similarly we can write the other blocks by taking the solutions of above equations as (0, 1) (1,0) and (1,1).

Given a block, how to find the interactions confounded?

The first step in detecting the interactions confounded in a replicate is to select the key block. If key block is not given, it is not difficult to obtain it. Select any treatment combination given in the given block and multiply all the treatment combinations in the given block by that treatment combination and we get the key block. From the key block we know the number of factors as well as the block size. Let it be n and k. We know then that the given design belongs to

the 2^n factorial in 2^r plots per block. The next step is to search out a unit matrix of order r. From these we can find the interaction confounded.

Example

Given the following block, find out the interactions confounded.

acde	ad
bcd	bde
e	ab
abec	c

Since the given block is not the key block. Let us multiply it by e. We get the following block.

acd	ade
bcde	bd
(1)	abe
abc	ce

This is the key block as it includes (1). It is obvious that the factorial involves five factors and has been confounded in 2^3 (=8) plots per block, Hence, the given design is $(2^5, 2^3)$.

	A	B	C	D	E
	1	0	1	1	0
	0	1	1	1	1
	0	0	0	0	0
	1	1	1	0	0
*	1	0	0	1	1
*	0	1	0	1	0
	1	1	0	0	1
*	0	0	1	0	1

* indicates the rows of a unit matrix of order 3.

A	B	C	D	E
1	0	0	$1(=(\alpha_1)$	$1(=\beta_1)$
0	1	0	$1 (=(\alpha_2)$	$0(=\beta_2)$
0	0	1	$0(=(\alpha_3)$	$1 (=\beta_3)$

The interaction confounded are $A^{\alpha_1}B^{\alpha_2}C^{\alpha_3}D$, $A^{\beta_1}B^{\beta_2}C^{\beta_3}E$. Here ABD and ACE are independent interactions confounded and BCDE is obtained as the product, of these two and is known as generalized interaction.

Confounding in 3^n Series

The concept of confounding here also is the same as in 2^n series. We shall illustrate the principles of confounding in 3^n in 3^r plots per block with the help of 3^3 experiments laid out in blocks of size $3^2(=9)$. Let the three factors be A, B and C and the confounded interaction be ABC^2. The three levels of each of the factor be denoted by 0, 1 and 2 and a particular treatment combination be $x_ix_jx_k$, i, j, k = 0, 1, 2.

Number of blocks per replication $=3^{n-r} = 3$

Block size $= 3^r = 9$

Degrees of freedom confounded =2

Number of interactions confounded per replicate

$$= \frac{3^{n-r}-1}{3-1} = 1$$

The number of treatments in 3 blocks are determined by solving the following equations mod (3)

$x_1+x_2+2x_3=0$

$x_1+x_2+2x_3=1$

$x_1+x_2+2x_3=2$

Replication I

Block I			Block II			Block III		
A	B	C	A	B	C	A	B	C
1	0	1	1	0	0	1	0	2
0	1	1	0	1	0	0	1	2
1	1	2	1	1	1	1	1	0
2	0	2	2	0	1	2	0	0
0	2	2	0	2	1	0	2	0
2	1	0	2	1	2	2	1	1
1	2	0	1	2	2	1	2	1
2	2	1	2	2	0	2	2	2
0	0	0	0	0	2	0	0	1

Balanced Design

A partially confounded design is said to be balanced if all the interaction of particular order are confounded in equal number of replication.

How to construct a Balanced Factorial Design

Example: Construct a $(2^5, 2^3)$ balanced design achieving balance over three and four factor interactions.

Solution: Total no. of treatment combination = $2^5 = 32$

Number of blocks per replicate= $2^{5-3} = 4$

The number of 3 factor interactions = $({}^5C_3) = 10$

The number of four factor interactions = $({}^5C_4) = 5$

So the total degrees of freedom to be confounded = 10+5 = 15

Since the design is $(2^5, 2^3)$, the degrees of freedom that can be confounded per replicate = $2^{5-3} - 1 = 3$.

So the number of replicates required = 15/3 = 5

Since the total number of three factor interactions is 10 and we have five replicates so each block confounds two, three factor interactions such that generalized interaction is of four factor.

Three factor interactions are ABC, ABD, ABE, ACD, ACE, ADE, BCD, BCE, BDE, CDE. Four factor interactions are ABCD, ABCE, ABDE, ACDE, BCDE.

Confounding ABD and ACE using equations

$X_1 + X_2 + X_4 = 0, 1$

$X_1 + X_3 + X_5 = 0, 1$

would result in replication I consisting of four blocks. Here only the key block each of the replication is given. Similarly Replication II confounds ACD, BCE a ABDE, Replication III confounds BCD, ABCE and ADE, Replication IV confounds ABCD, ABE and CDE, Replication V confounds ABC, BDE and ACDE.

Replication I

A	B	C	D	E	
1	0	0	1	1	ade
0	1	0	1	0	bd
0	0	1	0	1	ce
1	1	0	0	1	abe
1	0	1	1	0	acd
0	1	1	1	1	bcde
1	1	1	0	0	abc
0	0	0	0	0	(1)

Replication II

A	B	C	D	E	
1	0	0	1	0	ad
0	1	0	0	1	be
0	0	1	1	1	cde
1	1	0	1	0	abd
1	0	1	0	1	ace
0	1	1	1	0	bcd
1	1	1	0	0	abc
0	0	0	0	0	(1)

Replication III

A	B	C	D	E	
1	0	0	0	1	ae
0	1	0	1	1	bde
0	0	1	1	1	cde
1	1	0	1	0	abd
1	0	1	1	0	acd
0	1	1	0	0	bc
I	1	1	0	1	abce
0	0	0	0	0	(1)

Replication IV

A	B	C	D	E	
1	0	0	1	1	ade
0	1	0	1	1	bde
0	0	1	1	0	cd
I	1	0	0	0	ab
1	0	1	0	1	ace
0	1	1	0	1	bce
1	1	1	1	0	abcd
0	0	0	0	0	(1)

Replication V

A	B	C	D	E	
1	0	0	1	0	ac
0	1	0	1	1	bce
0	0	1	0	1	de
1	1	0	0	1	abe
1	0	1	1	1	abce
0	1	1	1	0	bdc
1	1	1	0	0	abd
0	0	0	0	0	(1)

❐ ❐ ❐

Chapter – 10

Split and Strip Plot Design

The most common procedure for experiments with factorial sets of treatments is to use a randomized block design and assign the treatment combinations at random to plots within the blocks. However the requirement that all combinations be assigned at random can lead to another difficulty for conducting the experiment. For example

1. An agronomist might be interested in tillage requirements as one factor and nitrogen level as another. Because tillage treatments require machinery, they require larger plots, while nitrogen can be applied by hand and can be applied to smaller plots.
2. Experiments involving planting date as a factor would be easier to conduct if all the plots to be planted at one time could be grouped together.

These examples illustrate the type of treatments that require different sizes of plots for their efficient application. So it would be convenient to have an experimental design in which level of one factor could be applied to larger plots while the levels of another factor are applied to smaller plots. This is possible with split plot design.

In the split plot design the level of one factor are assigned at random to larger plots and the larger plots are then divided into small plots within the large plots. The larger units are called whole plots or main plots, while the smaller units are called split plots or subplots.

For example the main plot factor are three methods of tillage: T_1,

T_2, and T_3 and subplots factor might be four barley varieties: V_1, V_2, V_3 and V_4. The field plan for two replications will be

Rep-I					Rep - II		
T_3	T_2	T_1			T_1	T_3	T_2
V_3	V_4	V_2			V_1	V_2	V_3
V_1	V_1	V_4			V_3	V_1	V_4
V_2	V_3	V_3			V_2	V_3	V_1
V_4	V_2	V_1			V_4	V_4	V_2

Here the precision for the measurement of the effects of the main plot factor is sacrificed to improve that of subplot factor. Main plot treatment is estimated with low precision while subplot treatment and the interaction are estimated with higher precision.

Advantages

1. It permits the efficient use of some factors that require different sizes of plots for their applications.
2. It permits the introduction of new treatments into an experiment that is already in progress.
3. It provides increased precision in the estimation of some of the factorial effects.

Disadvantages

1. The whole plot factor is estimated with less precision than the sub plot. Hence large differences are often required for significance with the whole plot factor, while the subplot factor might have significant effects that are too small to be of practical significance.
2. Statistical analysis is more complex than that required for the RBD because different standard errors are required for different comparisons.

Model

Let A be the main treatment with p levels, a_1, a_2 .....a_p and B be

sub treatment with q level, b_1, b_2.....b_q. Let r be the number of replications.

Further, let Y_{ijk} denote the observation of a_jb_k treatment combination in i^{th} replicate. Then the model is

$$Y_{ijk} = \mu + r_i + a_j + e_{ij} + b_k + (ab)_{jk} + e_{ijk}$$

$i = 1, 2, ...r$ $\quad j = 1, 2, ... p,$ $\quad k = 1, 2, ... q$

where μ is the general mean, r_i, a_j, and b_k are the effects of i^{th} replication, j^{th} level of A, k^{th} level of B and $(ab)_{jk}$ is the interaction effect, e_{ij} and e_{ijk} are the random errors distributed independently and normally with mean o and variances σ_a^2 and σ_b^2 respectively i.e. $e_{ij} \sim N(0, \sigma_a^2)$ and $e_{ijk} \sim N(0, \sigma_b^2)$

Method of Analysis

Let A be the main treatment with a levels as a_1, a_2 ....., a_p and B be sub treatment with q levels as b_1, b_2,....., b_q. Let r be the number of replications. The data are entered in the following date sheet. From the marginal totals and sub totals prepare R x A and A x B tables.

Replicate					
Treatment combination	1	2 . . i . .r			Totals
a_1b_1 a_1b_2 .. a_1b_q					a_1b_1 a_1b_2 ... a_1b_q
Sub total	r_1a_1	r_2a_1		r_ia_i r_ra_1	
a_2b_1 a_2b_2 .. a_2b_q					a_2b_1 a_2b_2 ... a_2b_q

Sub total	r_1a_2	r_2a_2	r_ia_2	r_ra_2	
a_pb_1					a_pb_1
a_pb_2					a_pb_2
					...
a_Pb_q					a_pb_q
Sub total	r_1a_P	r_2a_p	r_ia_p	r_ra_p	

R x A table

	r_1	r_2 . . . . . r_r	Totals
a_1	r_1a_1	r_2a_1 . . . . . r_ra_1	A_1
a_2			A_2
.			
a_p	r_1a_p	r_2a_p . . . . . r_ra_p	A_p
Totals	R_1	R_2 . . . . . R_r	G

A x B Table

	a_1	a_2 . . . a_p	Total
b_1	a_1b_1	a_2b_1 . . . a_pb_1	B_1
b_2	a_1b_2	a_2b_2 . . . a_pb_2	B_2
.	.	. . . .	
b_q	a_1b_q	a_2b_q . . . a_pb_q	B_q
Totals	A_1	A_2 . . A_p	G

Let $R_1, R_2, \ldots\ldots\ldots\ldots R_r$ be totals of r replicates

$A_1, A_2 \ldots\ldots\ldots\ldots A_p$ be totals of a levels of factor A

$B_1, B_2 \ldots\ldots\ldots\ldots B_q$ be totals of b levels of factor B

Then $N = rpq$, $G = \Sigma\Sigma\Sigma Y_{ijk} = \Sigma R_1 = \Sigma A_j = \Sigma B_k$ = Grand Total

C.F. $= G^2/N$

ANOVA Table

Source	d.f.	S.S.	M.S.	F
Replicates	r-1	$\sum R_i^2/pq$-C.F. = SSR	MSR	
Main treat. (A)	p-1	$\sum A_i^2$/ rq-C.F.=SSA	MSA	MSA/E_a
Error(a)	(r-1)(p-1)	$\sum\sum(r_ia_j)^2/q$ – C.F.-SSR-SSA=SSE_a	E_a	
Sub treat. (B)	q-1	$\sum B_k^2/rp$ – C.F. =SSB	MSB	MSB/E_b
Main x Sub	(p-1) (q-1)	$\sum\sum(a_jb_k)^2/r$-C.F.-SSA-SSB=SSAB	MSAB	$MSAB/E_b$
Error (b)	p(q-1) (r-1)	By subtraction = SSE_b	MSE_b	
Total	rpq-1	$\sum Y^2$-C.F.		

The main treatment A is tested against error (a) while sub treatment and the interaction A x B are tested against Error (b).

Standard Errors and Critical Differences

The standard errors for the comparison of the different treatment means are presented in the form of the table.

Difference between	**Notation**	**SE_d**	**C.D. at α level of significance**
Two A mean	$\bar{a}_i - \bar{a}_j$	$\sqrt{2E_a/rq}$	$SE_{(d)} \times t_{[\alpha,\ (r-1)(p-1)]}$
Two B mean	$\bar{b}_i - \bar{b}_i$	$\sqrt{2E_b/rq}$	$SE_{(d)} \times t_{[\alpha,\ p(r-1)(r-1)]}$
Two B means at the same level A	$\overline{a_ib_j} - \overline{a_ib_k}$	$\sqrt{2E_b/r}$	$SE_{(d)} \times t_{[\alpha,\ p(r-1)(r-1)]}$
Two B means at the same or different level of A	$\overline{a_ib_j} - \overline{a_jb_j}$ or $\overline{a_ib_j} - \overline{a_jb_k}$	$\sqrt{2[(q-1)E_b + E_a]/rq}$	$SE_{(d)} \times t_{[\alpha,\ p(r-1)(r-1)]}$

In the case of treatment means of two main treatments at the same level of sub-treatments or different levels of sub-treatment, the ratio of difference between two treatment means and its SE does not follow students t-distribution. Let t_1 and t_2 be tabulated value of t corresponding to the degree of freedom for main plot error and sub plot error, then the tabulated value of t for the above said comparison becomes.

$$t = \frac{(q-1)E_b t_2 + E_a t_1}{(q-1)E_b + E_a}$$

Presentation of results:

A x B mean table (which is obtained by dividing A x B total table contents by r and taking means on both sides) along with various standard errors and C.D. values can be presented as follows:

A x B Mean Table					
	a_1	a_2	. .	a_p	Means
b_1	a_1b_1/r	$a_2b_1/_r$	. . .	a_pb_1/r	b_1
b_2	a_1b_2/r	$a_2b_2/_r$	. .	a_pb_2/r	b_2
.	.	.	. . .	.	
b_q	a_1b_q/r	a_2b_q/r	. . .	a_pb_q/r	b_q
Means	a_1	a_2	.	a_p	G

SE_d for A = B = A x B =

C.D. for A =; B= ; Ax B =

Spilt-split plot design

The sub plots in spilt plot design can further be divided into sub plots to accommodate one more factor. For example, irrigation levels will be taken as main treatments, levels of nitrogen as sub-treatment and methods of application of nitrogen as sub sub-treatment. The method of analysis of this design will have three error components instead of two. Let the number of replications is 'r' main plot

treatments 'p' sub plot treatment 'q' and sub sub-treatment 's'. It is a simple extension of split plot design and hence sub-sub treatments are more precisely compared than sub treatments which are in turn more precisely compared than main treatments. The different sum of squares is computed on the basis of sub-sub unit. The divisor at each stage will be the number of sub-sub units involved in the numerator.

ANOVA Table

Source	d.f.	SS	M.S.	F_{cal}
Replication	(r-1)	SSR	MSR	$MSR/MSE_1 \sim F_{(r-1),\,(r-1)(p-1)}$
A	(p-1)	SSA	MSA	$MSA/MSE_1 \sim F_{(p-1),\,(r-1)\,(p-1)}$
Main plot error (E_1)	(r-1)(p-1)	SSE_1	MSE_1	
Sub plot treatment B	(q-1)	SSB	MSB	$MSB/MSE_2 \sim F_{(q-1),\,p(q-1)\,(r-1)}$
Interaction AB	(p-1) (q-1)	SSAB	MSAB	$MSAB/MSE_2 \sim F_{(p-1)\,(q-1),\,p(q-1)\,(r-1)}$
Sub plot error (E_2)	p(q-1)(r-1)	SSE_2	MSE_2	
Sub Sub treatment C	(s-1)	SSC	MSC	$MSC/MSE_3 \sim F_{(s-1),\,pq(r-1)\,(s-1)}$
Interaction AC	(p-1) (s-1)	SSAC	MS/AC	$MSAC/MSE_3 \sim F_{(p-1)\,(s-1),\,pq(r-1)\,(s-1)}$
AB	(q-1) (s-1)	SSAB	MSAB	$MSAB/MSE_3 \sim F_{(q-1)\,(s-1),\,pq(r-1)\,(s-1)}$
ABC	(p-1) (q-1) (s-1)	SSABC	MSABC	$MSABC/MSE_3 \sim F_{(p-1)\,(q-1)\,(s-1),pq(r-1)\,(s-1)}$
Sub-sub plot error (E_3)	pq(r-1) (s-1)	SSE_3	MSE_3	
Total	rpqs-1			

The standard error for the difference of two treatments means are same for the main unit and sub plot treatments as the case of split plot design except the extra divisor $\sqrt{s}$. The standard errors for remaining comparison are given below.

Standard Errors

Difference between	Notation	SE_d
Two C Means	$\bar{c}_i - \bar{c}_j$	$\sqrt{2MSE_3/rpq}$
Two C means at the same level of A	$\overline{a_ic_j} - \overline{a_ic_k}$	$\sqrt{2MSE_3/rq}$
Two C means at the same level of B	$\overline{b_ic_j} - \overline{b_ic_k}$	$\sqrt{2MSE_3/rp}$
Two C means at the same level of AB	$\overline{a_ib_jc_k} - \overline{a_ib_jc_l}$	$\sqrt{2MSE_3/r}$
Two B means at the same or different level of C	$\overline{b_ic_k} - \overline{b_jc_k}$ or $\overline{b_ic_k} - \overline{b_ic_e}$	$\sqrt{2[(s-1)MSE_3 + MSE_2]/rps}$
Two BC means at the same level of A	$\overline{a_ib_ic_j} - \overline{a_ib_jc_k}$	$\sqrt{2[(s-1)MSE_3 + MSE_1]/rs}$
Two A means at same or different level of C	$\overline{a_ic_j} - \overline{a_jc_j}$ or $\overline{a_ic_j} - \overline{a_jc_k}$	$\sqrt{2[(s-1)MSE_3 + MSE_1]/rsq}$
Two A means at the same level of B and C	$\overline{a_ib_kc_l} - \overline{a_jb_kc_l}$	$\sqrt{2[(s-1)MSE_3 + (q-1)MSE_2 + MSE_1]/rsq}$

Missing Observation in Split Plot Design

Let a single sub unit in which the treatments 'a_ib_j' occurs is missing in split plot layout, then the estimate of the missing observation is

$$Y_{ij} = \frac{rA_i + q(a_ib_j) - (a_i)}{(r-1)(q-1)}$$

where A_i is the total of the whole or main unit in which missing sub unit occurs, (a_ib_j) be the total of all the remaining sub units which receive the sub treatment a_ib_j and (a_i) be the total of all the sub units which receive the whole unit, a_i over all replications. If one sub unit is missing one d.f. is subtracted from errors of sub plot and also from the total. The treatment sum of squares and error for main plots are slightly inflated due to substitution of missing value.

Experiments involving more than two factors

When number of factors are to be tested is more than two then either split split plot design may be used or two or more factors can be combined in main treatments or in sub treatments or both in main x treatments.

Illustration

Give layout, analyse and precision of various factors and their interactions of a split plot design with 3 replication and having varieties = 2 (V_1, V_2), fertilizers = 3 (F_1, F_2, F_3); Pesticides = 2 (P_1, P_2) when (i) fertilizers are taken as main plot and varieties and pesticides both are taken as subplot treatment (ii) fertilizers and pesticides are both taken as main treatments and varieties are taken as sub treatment.

Solution

Case I: Fertilizer are taken as main treatment while varieties and pesticides both are taken as sub treatments. The layout is

RI				RII				RIII		
F_2	F_1	F_3		F_1	F_3	F_2		F_3	F_1	F_2
V_2P_1	V_1P_1	V_2P_2		V_1P_1	V_2P_1	V_2P_2		V_2P_2	V_1P_1	V_2P_1
V_1P_2	V_2P_2	V_1P_2		V_2P_2	V_1P_1	V_1P_2		V_1P_1	V_1P_2	V_2P_2
V_1P_1	V_1P_2	V_1P_1		V_1P_2	V_2P_2	V_2P_2		V_1P_2	V_2P_1	V_1P_1
V_2P_2	V_2P_1	V_2P_1		V_2P_1	V_1P_2	V_1P_1		V_2P_1	V_2P_2	V_1P_2

From the treatments totals prepare the various interaction tables like P x F, F x V and V x P as explained in the procedure earlier and find totals for replicates, fertilizers, varieties and pesticides. Let R_1, R_2, R_3; F_1, F_2, F_3; V_1, V_2, V_3 and P_1, P_2 be replicate totals, fertilizer totals, varieties totals and pesticide totals respectively.

ANOVA Table

Source	d.f.	SS	M.S.	F_{cal}
Replicates	r-1 = 2	$\sum R_1^2/12 - C.F.$	MSR	MSR/ MSE_a
Main treatments				
(Fertilizers)	f-1 = 2	$\sum F_j^2/12 - C.F.$	MSF	MSF/ MSE_a
Error (a)	(r-1) (f-1)=4	$\frac{\sum\sum (r_1 f_j)^2}{4} - C.F. - SSR - SSF$	MSE_a	
Sub treatments				
(varieties)	(v-1) = 1	$\sum V_k^2/18 - C.F.$	MSV	MSV/MSE_b
Pesticides	p -1=1	$\sum P_i^2/18 - C.F.$	MSP	MSP/MSE_b
Var. x pesticides	(v-1)(p-1)=1	$\frac{\sum\sum (v_k p_i)^2}{9} - C.F. - SSV - SSP$	MSVP	$MSVp/MSE_b$
Main x sub:				
Fert. x varieties	(f-1) (v-1)=2	$\frac{\sum\sum (f_j v_k)^2}{9} - C.F. - SSF - SSV$	MSFV	$MSFV/MSE_b$
Fert. x pesticides	(f-1) (p-1) =2	$\frac{\sum\sum (f_j p_i)^2}{6} - C.F. - SSF - SSP$	MSFP	$MSFP/MSE_b$
Fert. x varieties x pesticides	(f-1)(v-1) (p-1)=2	$\frac{\sum\sum\sum (f_j v_k p_i)^2}{3} - C.F. - SSV - SSF - SSP - SSFP - SSVF - SSVP$	MSFVP	$MSFVP/MSE_b$
Error (b)	(-)	By subtraction	MSE_b	
Total	rfvp-1=35	$\Sigma Y^2 - C.F.$		

Here the main treatment (fertilizer) is estimated with low precision while both the sub treatments (varieties and pesticides) and

all the interaction are estimated with higher precision.

Presentation of results: Prepare the following mean tables in cyclic order from tables of totals and take means, give the standard error of mean or difference and CD values.

P x F					F x V				V x P		
	F_1	F_2	F_3			V_1	V_2			P_1	P_2
P_1					F_2				V1		
					F_2						
P_2					F_3				V_2		
Means	f_1	f_2	f_3		Means	v_1	v_2		Means	p_1	p_2

Case-II: When combinations of fertilizer and pesticides are taken as main plot treatments and varieties as sub-treatments.

Layout

R_1	F_1P_2	F_2P_1	F_3P_1	F_2P_2	F_3P_1	F_1P_1
	V_2	V_1	V_2	V_1	V_2	V_1
	V_1	V_2	V_1	V_2	V_1	V_2

R_2	F_1P_1	F_2P_2	F_3P_1	F_1P_2	F_3P_2	F_2P_1
	V_2	V_1	V_2	V_1	V_2	V_1
	V_1	V_2	V_1	V_2	V_1	V_2

R_3	F_3P_2	F_2P_2	F_2P_4	F_1P_1	F_1P_2	F_3P_1
	V_1	V_2	V_1	V_2	V_2	V_2
	V_2	V_1	V_2	V_1	V_1	V_1

Prepare P x F, F x V and V x P tables of totals as explained in the previous case and find totals for various levels of fertilizer, pesticide and varieties. Let R_1, R_2, R_3; F_2, F_3 ; V_1, V_2, V_3 and P_1, P_2 be replicate totals, fertilizer totals, varieties totals and pesticides totals respectively.

ANOVA Table

Source	d.f.	S.S.	M.S.	$F_{cal.}$
Replicates	r-1=2	$\Sigma R_i^2/12 - C.F.$	R	
Main treatment				
(fertilizer)	f-1 =2	$\Sigma F_j^2/12 - C.F.$	F	F/E_a
Pesticides	p-1 = 2	$\Sigma P_i^2/18 - C.F.$	P	P/E_a
Fert. x pesticides	(f-1) (p-1)=2	$\frac{\Sigma\Sigma(f_j p_i)^2}{6} - C.F. - SSF - SSP$	FP	FP/E_a
Error(a)	(r-1) (fp-1)=1	$\frac{\Sigma\Sigma(r_i f_j p_i)^2}{2} - C.F. - SSR - SSF - SSP - SSFP$	E_a	
Sub-treatment				
Varieties	(v-1) =1	$\Sigma V_k^2/18 - C.F.$	V	V/E_b
Main x sub				
Fert. x varieties	(f-1) (v-1) =2	$\frac{\Sigma\Sigma(f_j v_k)^2}{6} - C.F. - SSF - SSV$	FV	FV/E_b
Fert. x pesticides	(f-1) (p-1) = 2	$\frac{\Sigma\Sigma(f_j p_i)^2}{6} - C.F. - SSF - SSP - SS\,FP$	FP	FP/E_b
Fert. x varieties x pesticides	(f-1) (p-1) (v-1) =2	$\frac{\Sigma\Sigma(f_j p_i v_k)^2}{6} - C.F. - SSF - SSP - SS\,FP - SSV - SSFV - SSV\,P$	FVP	FVP/E_b
Error(b)	(-)	(-)	E_b	
Total	rfvp-1=35	$\Sigma Y^2 - C.F.$		

Here both the main treatments fertilizers and pesticides and their interaction FxP are estimated with low precision, while the sub treatment varieties and all other interaction of various factors with varieties are estimated with higher precision. The results are presented in the same way as in the previous case.

Example:

A wheat breeder wanted to determine the effect of planting date on the yield of four varieties of winter wheat.

Planting date: D_1 = Oct. 10; D_2 = Oct. 20; D_3 = Oct. 30

Varieties are: V_1; V_2; V_3; V_4

To facilitate the operation of planting and harvesting a split-plot design with planting dates assigned at random to main plots within each of three blocks. Varieties were then assigned at random to plots with planting dates. The field plan and yields (kg/plot) are given below. Analyse the data and describe results.

Blocks I

Dates	D_2	D_1	D_3
	V_2	V_1	V_1
	26	27	19
	V_3	V_4	V_3
	21	16	18
	V_1	V_2	V_4
	37	26	12
	V_4	V_3	V_2
	20	28	23

II

D_1	D_2	D_3
V_4	V_2	V_4
16	25	29
V_1	V_3	V_1
35	23	26
V_3	V_1	V_2
27	40	16
V_2	V_4	V_3
19	18	25

III

D_1	D_3	D_2
V_1	V_4	V_3
30	11	22
V_3	V_3	V_1
21	19	35
V_2	V_1	V_2
20	24	30
V_4	V_2	V_4
16	23	28

Data Analysis

The preliminary steps in the data analysis involve the construction of two tables of totals. These are shown as table 1 and 2. The analysis of variance of the yield data is presented in table 3.

$$C.F. = \frac{G^2}{rpq} = \frac{(851)^2}{36} = 20,116.69$$

$$\text{Total SS (SSTot)} = \sum_i \sum_j \sum_k y_{ijk}^2 - CF = 26^2 + \ldots\ldots + 28^2 - 20,116.69$$

$$= 1596.31$$

SS due to replication/ blocks (SSR)

$$= \frac{1}{pq}\sum_i R_i^2 - CF = \frac{1}{12}(273^2 + 299^2 + 279^2) - C.F. = 30.89$$

Table-1: Two-way table of Block x Date totals

	Block			
Date	I	II	II	**Sum**
D_1	97	97	87	**281**
D_2	104	106	115	**325**
D_3	72	96	77	**245**
Sum	**273**	**299**	**279**	**851**

Table-2: Two-way table of Variety x Date (totals)

	Variety				
Date	V_1	V_2	V_3	V_4	**Sum**
D_1	92	65	76	48	**281**
D_2	112	81	66	66	**325**
D_3	69	62	62	52	**245**
Sum	**273**	**208**	**204**	**166**	**851**

SS due to main plots (SSA)

$$= \frac{1}{rq}\sum_j A_j^2 - CF = \frac{1}{12}(281^2 + 325^2 + 245^2) - 20116.69 = 267.55$$

SS due to error A (SSEA) $= \frac{1}{q}\sum_i\sum_j R_{ij}^2 - CF - SSR - SSA$

$$= \frac{1}{4}(97^2 + \ldots + 77^2) - 20116.69 - 30.89 - 267.55 = 83.11$$

SS due to subplot (SSB)

$$= \frac{1}{ra}\sum_k R_k^2 - CF = \frac{1}{9}(273^2 + 208^2 + 204^2 + 116^2) - CF = 657.19$$

SS due to Interaction (SSAB) $= \frac{1}{r}\sum_j\sum_k T_{jk}^2 - CF - SSA - SSB$

$$= \frac{1}{3}(92^2 + \ldots + 52^2) - CF - 267.55 - 657.19 = 204.89$$

SS due to error (b) = SSTot – SSR – SSA – SSEA – SSB-SSAB

$$= 1589.31 - 30.89 - 267.55 - 83.11 - 657.19 - 204.89$$

$$= 352.67$$

Analysis of variance of yields (kg/plot) from a Variety X Planting dates trial with Winter Wheat

ANOVA				
Source	df	SS	MS	F
Block	2	30.89	15.44	0.74
Date	2	267.55	133.78	6.44
Error(a)	4	83.11	20.78	
Variety	3	657.19	219.06	11.18**
Variety x date	6	204.89	34.15	1.74
Error (b)	18	352.67	19.59	
Total	35	1596.306	45.61	

** significant at 1%.

Testing Significance

1. Tabulated 1% F at 2, 4 df = 18.00 > 6.44. Thus the main effect of the planting date is non significant.
2. Tabulated 1% F at 3,18 df = 5.09 < 11.18. Hence the main effect of the variety is highly significant.
3. Tabulated 5% F at 6, 18 df = 2.66 and the 1% F at 6, 18 df = 4.01.

   Since 2.66 > 1.74 and also 4.01 > 1.74, the date x variety interaction is not significant at the 5% level as well as at the 1% level.

Standard errors:

Means:

1. Planting date means: $SE = \sqrt{\frac{MSE_a}{rq}} = \sqrt{\frac{20.78}{12}} = 1.31$
2. Variety means: $SE = \sqrt{\frac{MSE_b}{rp}} = \sqrt{\frac{19.59}{9}} = 1.47$

3. Date x variety means: $SE = \sqrt{\frac{MSE_b}{r}} = \sqrt{\frac{19.59}{3}} = 2.55$

Differences:

1. Two date means: $SE_d = \sqrt{\frac{2MSE_a}{rq}} = \sqrt{\frac{(2)(20.78)}{12}} = 1.86$

2. Two variety means: $SE_d = \sqrt{\frac{2MSE_b}{rp}} = \sqrt{\frac{(2)(19.59)}{9}} = 2.08$

3. Two variety means on same date: $SE_d = \sqrt{\frac{2MSE_b}{r}} = \sqrt{\frac{(2)(19.59)}{3}} = 3.61$

4. Two date means for same or different variety:

$$SE_d = \sqrt{\frac{2[(q-1)MSE_b + MSE_a]}{rq}} = \sqrt{\frac{2[(3)(19.59) + 20.78]}{12}} = 3.64$$

Further CD value can be calculated

The results of an experiment to examine the effect of planting date on the yield of four varieties of winter wheat are summarized in Table 4.

Table 4: Mean yields (kg/plot) of four winter wheat varieties planted on three different dates

	Variety				
Date	1	2	3	4	**Mean**
Oct. 10	30.67	21.67	23.33	16.00	**23.42**
Oct. 20	37.33	27.00	22.00	22.00	**27.08**
Oct. 30	23.00	20.67	20.67	17.33	**20.42**
Mean	**30.33**	**23.11**	**22.67**	**18.44**	**23.64**

SE: date = 1.31, variety = 1.47; variety x date = 2.55

Strip Plot Design

Consider an experiment with methods of seed-bed preparation at

three levels: S_1, S_2 and S_3, and a nitrogen fertilizer factor of four levels: N_1, N_2, N_3 and N_4. Suppose that we wanted to use machinery to apply the nitrogen as well as to do the seed-bed preparation. In this case it would be convenient to use large plots for both factors. This is possible with the strip-plot design.

In a strip-plot design the levels of one factor are assigned to strips of plots running through the block in one direction. A separate randomization is used in each block. The levels of the second factor are then applied to strips of plots that are oriented perpendicularly to the strips for the first factor. The field plan for two replications of a seed bed x nitrogen experiment run as a strip-plot design might appear as shown below.

Replication I				Replication II			
	S_3	S_1	S_2		S_1	S_3	S_2
N_1				N_2			
N_2				N_3			
N_0				N_1			
N_3				N_0			

Field plan for strip-plot design for a seed bed x nitrogen experiment

In the strip plot design there are three sizes of plots:

1. The strip of four plots on which the seed-bed factor is applied.
2. The strip of three plots on which the levels of nitrogen are applied.
3. The small plots containing the NS combinations.

Because there are three size of plot, there are three experimental errors, one for each plot size. The interaction is measured with greater precision than are the main effects.

Advantages, Disadvantages, and Uses:

It permits the efficient application of factors that would be difficult to apply to small plots. Here both the factors are estimated with low precision while the interaction A x B is estimated with higher precision.

The disadvantage of the strip-plot design are similar to those of the split-plot design:

- Differential precision in the estimation of interaction and the main effects.
- Complicated statistical analysis.

The use of the strip-plot design are similar to the uses of the split-plot design:

- For experiments involving factors that are difficult to apply to small plots.
- To introduce new factors into an experiment already in progress.

Model

Let Y_{ijk} be the observations of the plot belonging to the i^{th} replicate, j^{th} levels of factor A and k^{th} level of second factor B i.e. to the treatment combination (a_jb_k) in the i^{th} replicate.

$$Y_{ijk} = \mu + r_i + a_j + e_{ij} + b_k + e_{ik} + (ab)_{jk} + e_{ijk}$$

$i = 1, 2, \ldots\ldots r, \quad j = 1, 2, \ldots\ldots p, \quad k = 1, 2, \ldots\ldots q.$

μ = general mean, $r_i = i^{th}$ block effects

a_j = effect of j^{th} level of factor A b_k = effect of k^{th} level of factor B

$(ab)_{jk}$ = Interaction between j^{th} level of factor A and k^{th} level of factor B

The error components e_{ij}, e_{ik} and e_{ijk} are independently and normally distributed with mean zero and respective variances σ^2_a, σ^2_b and σ^2_c

Prepare R x A, R x B and A x B tables of totals from the data and find totals of various replicates and various levels of factors A and B.

R x A				R x B				A x B				
	r_1	r_2.....r_r			r_1	r_2.....r_r			a_1	a_2.....a_p		Total
a_1				b_1				b_1				B_1
a_2				b_2				b_2				B_2
.				.				.				.
.				.				.				.
a_p				b_q				b_q				B_q
Total	R_1	R_2	R_r	Total	R_1	R_2	R_r	Total	A_1	A_2	A_p	G

Let R_1 R_2....., R_r be the totals of r replicates

A_1 A_2....., A_p be the totals of p levels of factor A

B_1 B_2....., B_q be the totals of q levels of factor B

Then N = rpq = Total No. of observations

$\Sigma\Sigma Y_{ijk} = \Sigma R_i = \Sigma A_j = \Sigma B_k = G$ = grand total

C.F. $= G^2/N$

ANOVA Table

Source	**d.f.**	**S.S.**	**M.S.**	**F**
Replicates	r-1	$\Sigma R_i^2/pq$-C.F.	R	
Factor A	p-1	$\Sigma A_j^2/rq$-C.F.	A	A/E_a
Error (a)	(r-1) (p-1)	$\Sigma\Sigma(r_i\ a_j)^2/q$- C.F. SSR-SSA	E_a	
Factor B	q-1	$\Sigma B_k^2/rp$-C.F.	B	B/E_b
Error (b)	(r-1) (q-1)	$\Sigma\Sigma(r_i\ b_k)^2/p$ - C.F. SSR-SSB	E_b	
Interaction A x B	(p-1) (q-1)	$\Sigma\Sigma(a_j\ b_k)^2/r$-C.F. SSA-SSB	AB	AB/E_c
Error (c)	(-)	by subtraction	E_c	
Total	rpq-1	$\Sigma\Sigma\Sigma y_{ijk}^2$-C.F.		

Standard error of difference and CD values for various

comparisons is obtained as follows:

Comparison	S.E. of difference	5% C.D.
Between two A means say $\bar{a}_1 - \bar{a}_2$	$\sqrt{2E_a/rq}$	$\sqrt{2E_a/rq}$ x $t_{[(\alpha, (r-1)(p-1)]}$
Between two B means say $\bar{b}_1 - \bar{b}_2$	$\sqrt{2E_b/rp}$	$\sqrt{2E_b/rp}$ x $t_{[(\alpha, (r-1)(q-1)]}$
Between two A means at same level of B say $\overline{a_1b_1} - \overline{a_2b_1}$	$\sqrt{2[(q-1)E_c + E_a]/rq}$	SE_d x t_α
Between two B means at same level of A say $\overline{a_1b_1} - \overline{a_1b_2}$	$\sqrt{2[(p-1)E_c + E_b]/rp}$	SE_d x t_β

$$\text{Here } t_\alpha = \frac{(q-1)E_c t_c + E_a t_a}{(q-1)E_c + E_a}$$

where t_a and t_c are the t values at error (a) errors (c) degree of freedom.

$$t_\beta = \frac{(p-1)E_c t_c + E_b t_b}{(p-1)E_c + E_b}$$

where t_b is t values at error (b) degree of treatments.

Remarks:

Like wise in Split plot design, in Strip plot design too, when the number of factors to be tested is more than two then two or more factors can be clubbed in factor A or in factor B or in both the f actors A and B as per the situation.

Example

With a view to formulate optimum spacing schedule for different varieties of wheat an experiment was conducted in a strip plot design.

The treatment were:

Varieties V_1, V_2, V_3, V_4, V_5 Spacings S_1, S_2, S_3, S_4

Numbers of replications = 3

The layout and the yield in kg/plot is given below:

	REPLICATION I					REPLICATION II			
	S_1	S_2	S_3	S_4		S_3	S_1	S_4	S_2
V_2	5.6	2.3	6.7	4.93	V_4	3.5	6.45	4.8	6.9
V_5	5.46	5.87	2.63	6.78	V_1	6.5	4.69	1.59	4.96
V_3	2.24	5.67	3.48	6.58	V_5	5.32	6.89	2.45	5.36
V_1	5.67	6.89	2.56	3.78	V_2	4.25	3.45	5.69	4.62
V_4	2.60	5.65	3.26	2.57	V_3	2.86	4.39	4.68	2.9

	REPLICATION III			
	S_1	S_2	S_3	S_4
V_3	4.26	6.89	4.35	2.89
V_4	5.69	4.89	4.58	5.36
V_5	2.56	6.89	3.25	4.60
V_2	8.90	2.68	4.89	6.09
V_1	3.89	2.68	1.89	2.80

Analysis

We form two-way tables of var x Repl. spacing x Repl and variety x spacing

		Replications				
		1	2	3	**Total**	Mean
	1	18.90	17.74	11.26	**47.90**	15.97
	2	19.53	18.01	22.56	**60.10**	20.03
Variety	3	17.97	14.83	18.39	**51.19**	17.06
	4	14.08	21.65	20.52	**56.25**	18.75
	5	20.74	20.02	17.30	**58.06**	19.35
	Total	**91.22**	**92.25**	**90.03**	**273.50**	

Replications						
		1	2	3	Total	Mean
	1	21.57	25.87	25.3	**72.74**	24.25
	2	26.38	24.74	24.03	**75.15**	25.05
Spacing	3	18.63	22.43	18.96	**60.02**	20.01
	4	24.64	19.21	21.74	**65.59**	21.86
	Total	**91.22**	**92.25**	**90.03**	**273.5**	

Spacings						
		1	2	3	4	**Total**
	1	14.25	14.53	10.95	8.15	**47.9**
	2	17.95	9.60	15.84	16.71	**60.10**
Variety	3	10.89	15.46	10.69	14.15	**51.19**
	4	14.74	17.44	11.34	12.73	**56.25**
	5	14.91	18.12	11.20	13.83	**58.06**
	Total	**72.74**	**75.15**	**60.02**	**65.59**	**273.50**

$$C.F = \frac{(GT)^2}{pqr} = \frac{(273.5)^2}{60} = 1246.70$$

$$\text{Variety SS (VSS)} = \frac{\sum V_i^2}{rq} - C.F. = \frac{47.9^2 + \ldots\ldots + 58.06^2}{12} - C.F.$$

$$= 8.45$$

$$\text{Replication SS} = \frac{\sum R_j^2}{pq} - C.F. = 0.123$$

$$\text{Total SS (1)} = \frac{V_1R_1^2 + \ldots\ldots + V_5R_5^2}{q} - C.F.$$

$$= \frac{18.9^2 + \ldots\ldots + 17.3^2}{4} - C.F. = 31.48$$

SS due to error E_1 = Total SS (1) – Replication SS – Var SS = 22.91

$$\text{Spacing SS} = \frac{\sum S_k^2}{rp} - C.F. = \frac{72.74^2 + \ldots\ldots + 65.59^2}{15} - C.F. = 9.50$$

$$\text{Total SS (2)} = \frac{S_1R_1^2 + \ldots\ldots + S_5R_5^2}{5} - \text{C.F.}$$

SS due to error E_2 = Total SS(2)–Spacing SS – Replication SS = 7.36

$$\text{Total SS (3)} = \frac{S_1V_1^2 + S_2V_2^2 + \ldots\ldots + S_5V_5^2}{3} - \text{C.F.}$$

$$= \frac{14.25^2 + \ldots\ldots + 13.83^2}{3} - \text{C.F.} = 52.25$$

SS due to interaction, VS = Total SS(3) – Varieties SS – Spacing SS = 34.29

Total sum of squares = $\Sigma\Sigma\Sigma\, Y_{ijk}^2$ – C.F. = 158.83

Error sum of squares (E_3)

= Total – Replication SS – Variety SS – Error SS (E_1) – Spacing SS – Error SS (E_2) – (VxS) SS = 76.19

ANOVA				
Source	d.f	SS	MSS	F
Replication	2	0.12	0.06	0.02
Variety, V	4	8.45	2.11	0.74
Error E_1	8	22.91	2.86	
Spacing, S	3	9.50	3.17	2.58
Error, E_2	6	7.37	1.23	
Variety x Spacing, VxS	12	34.29	2.86	0.90
Error, E_3	24	76.19	3.17	
Total	59	158.83		

SE of difference between two variety means

$$SE(d) = \sqrt{\frac{2MSE_1}{rq}} = 0.690$$

SE of difference between two spacing means = $\sqrt{\frac{2MSE_2}{pr}} = 0.4045$

Standard error of difference between two spacing means at same level of variety

$$SE(d) = \sqrt{\frac{2(3 \times 3.17 + 2.86)}{12}} = 1.43$$

SE of difference between two varieties at the same level of spacing

$$= \sqrt{\frac{2(4 \times 3.17 + 1.23)}{15}} = 1.36$$

❐ ❐ ❐

Chapter – 11

Covariance Analysis

It often happens that some source of variation which can not be controlled by the design can be measured by taking additional observations on some characters, called concomitant variable (x) which is highly correlated with the character of interest (y). Consider the case of rice variety trial in which weed incidence is used as a concomitant variable or covariate with grain yield, the character of primary interest. Similarly, in wheat crop, we take tiller number as the concomitant varable. The covariance analysis can adjust the mean yield in each plot to a common level of weed incidence. With the adjustment, the variations in yield due to weed incidence is quantified and effectively separated from that due to varietal difference. Covariance analysis can be applied to any number of covariates and to any type of functional relationship between variables. Here we will take a case of single covariate whose relationship to the character of interest is linear. Analysis of covariance in fact is statistical technique and not a design through which the error variation could be reduced by increasing the efficiency of the design.

Assumptions

1) The effect of different factors i.e. treatments, groups and regression are additive.
2) The yield is distributed normally and independently.

3) The covariate X or the concomitant variable X is not affected by the treatments.

4) The concomitant variable should be highly correlated with variable under study.

Analysis of covariance of RBD data

Let Y_{ij} denote the yield of i^{th} treatment in the j^{th} replication. i = 1, 2, ... v and j = 1, 2, ... r and X_{ij} be the corresponding value of the variable X. Assuming that there is no interaction between treatments and plots. We have a 2-way classification with one pair of observations per cell.

Let us assume that

$$Y_{ij} = \mu + t_i + b_j + \beta (X_{ij} - \bar{X}) + e_{ij}$$

where

X_{ij} = observation on concomitant variable

μ = general mean effect

t_i = effect of the i^{th} treatment

b_j = effect of the j^{th} block

β = regression coefficient of Y and X

e_{ij} = are random error components independently and normally distributed with mean zero and variance σ_e^2

Let

$$\sum_j Y_{ij} = T_i; \sum_i Y_{ij} = B_j; \sum_j X_{ij} = T_{ix}; \sum_i X_{ij} = B_{jx}$$

$$\sum_i T_i = \sum_j B_j = G; \qquad \sum_i T_{ix} = \sum_j B_{jx} = G_x;$$

The analysis of covariance will be of the form

ANOCOVA

Source	df	S_{XX}
Blocks	(r-1)	$S_{XX} = \sum_j B_{jx}^2/k - G_x^2/rk = B_{xx}$
		$S_{XY} = \sum_j B_j B_{jx}/k - G_x G/rk = B_{xy}$
		$S_{YY} = \sum_j B_j^2/k - G^2/rk = B_{yy}$
Treatment	$(\nu - 1)$	$S_{XX} = \sum_i T_{ix}^2/r - G_x^2/rk = T_{xx}$
		$S_{XY} = \sum_i T_i T_{ix}/r - G_x G/rk = T_{xy}$
		$S_{YY} = \sum_i T_i^2/r - G^2/rk = T_{yy}$
Error	(r-1) (ν-1)	S_{XX} = By subtraction(=E_{xx})
		S_{XY} = By subtraction (= E_{xy})
		S_{YY} = By subtraction (= E_{yy})
Total	($r\nu$ -1)	$S_{XX} = \sum\sum X_{ij}^2 - \frac{G_X^2}{rk} = G_{xx}$
		$S_{XY} = \sum\sum X_{ij}\ Y_{ij} - \frac{G_x G}{rk} = G_{xy}$
		$S_{YY} = \sum\sum Y_{ij}^2 - \frac{G^2}{rk} = G_{yy}$
Treatment + Error	r(k-1)	$S_{XX} = T_{xx}+E_{xx}=E^1_{xx}$
		$S_{XY} = T_{xy}+E_{xy}=E^1_{xy}$
		$S_{YY} = T_{yy}+E_{yy}=E^1_{yy}$

Our Hypothesis is $H_0 : t_1 = t_2 = \ldots\ldots = t_k$

Again adjusted Error Mean Square, E, with [(r-1)(ν-1) – 1]d.f.

$= E_{yy} - \beta E_{xy}$ where $\beta = E_{xy}/E_{xx}$

$= E_{yy} - (E_{xy})^2/E_{xx}$

Adjusted Error Mean Square E_1 (under H_0), with [(ν-1)-1]d.f

$= E^1_{yy} - (E^1_{xy})^2/E^1_{xx}$

Adjusted Treatments ss = $E_1 - E$ with (v-1)d.f.

Then the Hypothesis, H_0 is tested by F test, where

$$F = \frac{\text{Adjusted Treatment Mean Square}}{\text{Adjusted Error Mean Square}}$$

$$= \frac{E_1 - E}{E} \text{ with } (v-1), \text{ and } [(r-1)(v-1)-1]\text{d.f.}$$

If $F_{cal} \geq F_{tab}$ with (v-1),[(r-1)(v-1)-1]d.f. at chosen level of significance, th null hypothesis is rejected otherwise not.

The adjusted mean value for i^{th} treatment

$$\bar{T}_i = \bar{Y}_i - \beta(\bar{X}_i - \bar{X}..)$$

where $\bar{Y}_i = \sum_j Y_{ij}/r, \quad \bar{X}_i = \sum_j X_{ij}/r, \quad \bar{X}.. = \sum_i \sum_j X_{ij}/rv$

The difference between two adjusted treatment means is given by

$$\bar{T}_i - \bar{T}_i' = \bar{Y}_i - \bar{Y}_i' - \hat{\beta}(\bar{X}_i - \bar{X}_i')$$

$$V(\bar{T}_i - \bar{T}_i') = \frac{2\sigma^2}{r} + \frac{\sigma^2}{E_{xx}}(\bar{X}_i - \bar{X}_i')^2$$

So the variance of the estimate of the differences between any pair of treatment is not a constant. To make the critical difference a constant for each pair of treatment we have Average variance of treatment difference.

$$V(\bar{T}_i - \bar{T}_i') = \frac{2S_e^2}{r}\left[1 + \frac{T_{xx}}{(v-1)E_{xx}}\right]$$

where s_e^2 is the adjusted error mean square E. If $s_e^2 < E_{yy}/(r\text{-}1\ (v\text{-}1)$, then we say that analysis covariance is useful.

The effective error mean square in case of analysis of covariance s_e^2 is.

$$s_e^{2\prime} = s_e^2 \left[1 + \frac{T_{xx}}{(v-1)E_{xx}}\right]$$

Per cent gain in efficiency due to application of analysis of covariance as compared to simple analysis is given by

$$\left[\frac{s^2}{s_e^{2\prime}} - 1\right] x\,100$$

where $s^2 = \frac{E_{yy}}{(r-1)(v-1)}$

Example

An experiment was conducted to study the effect of level of protein in ration of cows on milk production. The three levels of protein are A, B and C. The design used was RBD with five replications. Here Y is the total milk yield for experimental period and X is the total milk yield for first 15 pre-experimental days after calving.

Analyse the data by using information about X for increasing the precision and draw conclusions using analysis of covariance technique.

Milk yield (in kg) of experimental animals

	A		B		C	
Blocks	**X**	**Y**	**X**	**Y**	**X**	**Y**
1	190	2400	175	1880	173	2251
2	170	1850	188	2250	225	2574
3	175	2430	195	2455	190	2812
4	150	1650	140	1858	162	1594
5	120	2630	199	2776	144	2890
Total	**805**	**10960**	**897**	**11219**	**894**	**12121**
Mean	161	2192	179.4	2243.8	178.8	2424.2

Calculations

$T_{1x} = 805$ $T_1 = 10960$

$T_{2x} = 897$ $T_2 = 11219$

$T_{3x} = 894$ $T_3 = 12121$

$B_{1x} = 538$ $B_1 = 6531$

$B_{2x} = 583$ $B_2 = 6674$

$B_{3x} = 560$ $B_3 = 7697$

$B_{4x} = 452$ $B_4 = 5102$

$B_{5x} = 463$ $B_5 = 8296$

$$G_x = \sum_i T_{ix} = \Sigma B_{j_x} = 2596$$

$$G = \Sigma T_i = \Sigma B_j = 34300$$

Various sum of squares can be calculated as follows:

$$B_{xx} = \frac{1}{v}\sum_j B_{jx}^2 - \frac{G_x^2}{rv} = 453868.67 - 449281.06 = 4587.6$$

$$B_{yy} = \frac{1}{V}\sum_j B_j^2 - \frac{G^2}{rv} = 80431355.33 - 78432666.67$$

$$= 1998688.67$$

$$B_{xy} = \frac{1}{v}\sum_j B_j B_{j_x} - \frac{G.G_x}{rv} = 5954030.67 - 5936186.67$$

$$= 17844.00$$

$$T_{xx} = \frac{1}{r}\sum_i T_{ix}^2 - \frac{G_x^2}{rv} = 450374.00 - 449281.07 = 1092.93$$

$$T_{yy} = \frac{1}{r}\sum_i T_i^2 - \frac{G^2}{rv} = 78581240.40 - 78432666.67$$

$$= 148573.73$$

$$T_{xy} = \frac{1}{r}\sum T_i^2 T_{ix} - \frac{G.G_x}{rv} = 5944483.4 - 5936186.67 = 8296.73$$

$$G_{xx} = \Sigma\Sigma X_{ij}^2 - \frac{G_x^2}{rv} = 459354.00 - 449281.07 = 10072.93$$

$$G_{yy} = \Sigma\Sigma Y_{ij}^2 - \frac{G^2}{rv} = 81001822.00 - 78432666.67 = 2569155.33$$

$$G_{xy} = \sum_{ij} X_{ij}\, Y_{ij} - \frac{GG_x}{rv} = 5979360.00 - 5936186.67 = 43173.33$$

$E_{xx} = G_{xx} - B_{xx} - T_{xx} = 10072.93 - 4587.60 - 1092.93 = 4392.40$

$E_{yy} = G_{yy} - B_{yy} - T_{yy} = 259155.33 - 1998688.67 - 148573.73$

$= 421892.93$

$E_{xy} = G_{xy} - B_{xy} - T_{xy} = 43173.33 - 17844.00 - 8296.73 = 17032.6$

Now complete the Analysis of variance table given below

ANCOVA

Source	d.f.	S_{xx}	S_{xy}	S_{yy}
Blocks	5	4587.6 (B_{xx})	17844 (B_{xy})	198688.67 (B_{yy})
Treatment	2	1092.93 (T_{xx})	8296.73 (T_{xy})	148573.73 (T_{yy})
Error	10	4392.40 (E_{xx})	17032.6 (E_{xy})	421892.93 (E_{yy})
Total	17	10072.93	43173.33	2569155.33
Treatment +Error	12	5485.33 (E'_{xx})	25629.33 (E'_{xy})	570466.66 (E'_{yy})

Step-II calculation of adjusted error mean square

$$\hat{B} = \frac{E_{xy}}{E_{xx}} = \frac{17032.60}{4392.40} = 3.87$$

Adjusted Error mean square $E = E_{yy} - \frac{(E_{xy})^2}{E_{xx}}$

$$= 421892.93 - \frac{(17032.6)^2}{4392.40} = 355844.87$$

Adjusted Error mean square under the null hypothesis (with 11 degree of freedom)

$$E_1 = E^1_{yy} - \frac{(E^1_{xy})^2}{E^1_{xx}} = 570466.66 - \frac{(25329.33)^2}{5485.33} = 453504.69$$

Adjusted treatment SS = $E_i - E$ = 5485.33 – 355844.87

= 97659.81 (with 2 d.f.)

Adjusted Analysis of variance

Source of variation	d.f.	SS	MS	F
Treatment	2	97659.81	48829.9	1.235
Error	9	355844.87	39538.31	

Table value of F at 5% level of significance with 2 and 9 d.f. is 4.26. So we infer that treatments do not differ significantly among themselves

Step-III: Calculation of Adjusted treatment means.

The adjusted treatment means are given by

$$\bar{T}_i = \bar{y}_i - \hat{\beta}(\bar{x}_i - \bar{x}..)$$

$$\bar{T}_1 = 2192.00 - 3.87\ (161.00 - 173.07) = 2238.71$$

$$\bar{T}_2 = 2219.30 \qquad \bar{T}_3 = 2402.02$$

Now s_e^2 = Adjusted error mean square = 39538.31 and the effective error mean square

$$s_e^{2'} = s_e^2 \left[1 + \frac{T_{xx}}{(v-1)E_{xx}} \right]$$

$$= 39538.31 \left[1 + \frac{1092.93}{2x4392.40} \right] = 44457.33$$

$$SE_d = \sqrt{\frac{2s^2e'}{r}} = \sqrt{\frac{2x\ 44457.33}{6_{\ 3}}} = 121.73$$

If the F value is significant then C.D. value can be calculated in the usual way.

❐ ❐ ❐

Chapter – 12

Multi-Locational Experiments

The purpose of research carried out at experimental stations is to formulate recommendations for the practitioners which consists of a population quite extensive either in space or time or both. Therefore it becomes necessary to ensure that the results obtained from researches are valid for at least several seasons in the future and over a reasonably heterogeneous space. A single experiment will precisely furnish information about only one place where the experiment is conducted and about the season in which the experiment is conducted. It has thus become a common practice to repeat on experiment at different places or over a number of occasions to obtain valid recommendations taking in to account place to place variations over time or both. In such cases the object of interest in analyzing a set of trials is, to estimate the average responses to given treatments and to test the consistency of these at different places and in different seasons. The test of absence or presence of interactions helps us to know whether or not the response is consistent from place to place.

The result of analysis may belong to any one of the following four types.

(1) Experimental errors homogeneous and the interaction absent.

(2) Experimental errors are homogeneous and the interaction

present.

(3) The error heterogeneous and the interactions absent and

(4) The error heterogeneous and the interaction present.

Thus while analyzing the data we first determine whether the experimental errors are homogeneous or not.

Test of homogeneity of experimental errors and treatment comparisons

If the magnitude of the experimental errors could be assumed to be the same at each place, a simple arithmetic mean of the treatment means can be taken to provide an estimate of treatment effects over the tract represented by the places and a simple analysis of variance on two way table showing treatment and place means would give the required information.

It experimental error differ from place to place, the simple mean and analysis of variance can not be used. Therefore, it is necessary to examine first the homogeneity of error variances by using Bartlett's test.

The data are first analyzed in usual way for each place (or year) separately. Let there be p mean squares s_1^2, s_2^2, ...s_p^2 based on respectively n_1, n_2,..... n_p degree of freedom. Here p is the number of places or years on which experiment is performed. From these values a pooled estimate of variance $\bar{s}^2$ is calculated.

$$\bar{s}^2 = \frac{\sum_{i=1}^{p} n_i s_i^2}{\sum_{i=1}^{p} n_i}$$

Next homogeneity of error variance is tested using Bartlett's χ^2 test.

$$\chi^2_{p-1} = \frac{\sum_{i=1}^{p} n_i \log_e \frac{\bar{s}^2}{s_i^2}}{1 + \frac{1}{3(p-1)}\left[\sum_{i=1}^{p}\frac{1}{n_i} - \frac{1}{\sum_{i=1}^{p} n_i}\right]}$$

If $\chi^2_{p-1} > \chi^2$ table value at α percent level of significance at (p-1) d.f then it is concluded that variances are significantly heterogenous otherwise error variances are said to be homogenous.

If the error variances are homogenous, then pooled analysis is carried out. The analysis of variances has the following setup for p randomized block experiments in t treatments conducted at p sites.

ANOVA

Source	d.f	M.S.
Places	(p-1)	s^2_p
Treatments	(t-1)	s^2_t
Treatment x places	(t-1) (p-1)	s^2_i
Pooled error	sum of error d.f of all the places	s^2_r

The error obtained from individual trials are pooled to obtain a joint estimate of error variance s^2e.

Testing the significance of treatment x place interaction

In order to see whether treatment x place interaction is significant i.e. whether the differences between treatments tends to vary from place to place, the mean square for treatment x place is compared with the estimate of error variance by F test. If this mean square is found to be non significant the interaction is considered to be absent.

It interaction is assumed to be non-existent the SS for treatment x places and the SS for error can be pooled to obtain a more precise estimate of error for testing the significance of treatment and place differences. In case treatment x place interaction can not be taken as

zero or in other words if the interaction between treatment and place is found to be significant, then the appropriate mean square for testing the significance of treatment as well as place differences is the mean square due to treatment x place.

We can test the significance of treatment mean square by treatment x place interaction. Similarly significance of places means can also be tested but usually it has no importance. This way we can analyse the data where (i) the errors are homogenous and interaction significant (ii) errors are homogenous and interactions non significant.

When the error variances are hetrogeneous, the procedure to be followed for testing the treatment difference depends on the presence or absence of the treatment x place interaction. The next step to be followed therefore consists in making a test of significance for the interaction. This is done by method of weighted analysis, the weights for each experiment being calculated as

$w_i = r/s^2_i$. For these weights, we calculate for each case the quantities w_i, p_i and $\Sigma w_i t_{ij}$ where pi are the place total for each treatment and t_{ij} are means for j-th treatment at i-th place. The calculation are given in the following table.

Treatments	Places 1	2.....	i.....	p	$\Sigma w_i t_{ij}$
1	t_{11}	t_{21}.....	t_{i1}.....	t_{p1}	$\sum_i w_i t_{i1}$
2	t_{12}	t_{22}.....	t_{i2}.....	t_{p2}	$\sum_i w_i t_{i2}$
.					
j	t_{1j}	t_{2j}.....	t_{ij}.....	t_{pj}	$\sum_i w_i t_{ij}$
T	t_{1t}	t_{2t}.....	t_{it}.....	t_{pt}	$\sum_i w_i t_{it}$
$w_i = \frac{r}{s_i^2}$	w_1	w_2.....	w_i.....	w_p	
$p_i = \sum_j t_{ij}$	p_1	p_2.....	p_i.....	p_p	
$s_i = \sum_j t_{ij}^2$	s_1	s_2.....	s_i.....	s_p	

We then compute the following $C.F. = \frac{(G.T.)^2}{t\sum_i w_i}$ Where

$G.T. = \sum_j (\sum_i w_i t_{ij})$

$$\text{Total SS} = \sum_{i=1}^{p} w_i s_i - CF \; ; \; \text{Treatment SS} = \frac{\sum_j (\sum_i w_i t_{ij})^2}{\sum_i w_i} - CF$$

$$\text{SS due to place (or years)} = \frac{1}{t}\sum_{i=1}^{p} w_i p_i^2 - CF$$

SS due to interaction between treatments and places

= Total SS - Treatment SS – SS due to places.

For testing the significance of interaction, we transform the SS due to interaction in to χ^2 using the formula.

$$\chi^2 = \frac{(n-4)(n-2)}{n(n+t-3)} \text{ (SS due to interaction)}$$

$$\text{with d.f.} = \frac{(p-1)(t-1)(n-4)}{(n+t-3)}$$

Where n is the error degree of freedom on whlch the error mean square is based in each experiment. If the interaction is present set down the treatment means of each places in two-way table and carry out a simple analysis of variance and compare the treatment MS with interaction MS. The procedure is known as unweighted analysis. The simple arithmetic mean of treatment means here provide the estimate of the treatment effect over the tract. It is an approximate method

When the interaction is absent, no general test tor over all treatment differences appears to be available, but we can test the individual treatment means by the following procedure.

From each trial set down the responses correspondingly to the treatment means in which we may be interested. The following weighted mean is then computed.

$\bar{y}_w = \frac{\sum_i w_i' y_i}{\sum_i w_i'}$ where $w_i' = \frac{1}{v(y_i)}$ and y_i is the response for i^{th} trial

χ^2 is now given as

$$\chi^2_{(p-1)} = (p-1) + \sqrt{\frac{(n-4)}{(n-1)}\left(\frac{(n-1)}{n}\theta - (p-1)\right)}$$

where $\theta = \sum_i w_i' y_i^2 - \bar{y}_w \sum_i w_i' y_i$

It χ^2 is significant we test the treatments against interaction as in the case of unweighted analysis. However if χ^2 is non-significant, then we compute the statistic.

$t = \bar{y}_w \Big/ \sqrt{\frac{1}{\sum_i w_i'}}$ which follows the t-distribution with p(n-1) degree of freedom.

Example: A set of trials was conducted at four locations in U.P. to compare four treatments. At each place randomized block design was used to carry out the experiment. The analysis of each trial individually provided an error mean square based on 12 degrees of freedom. The total yield in kg for each treatment over five replicates at each place as well as error mean squares estimated in each trial are given below.

	Blocks			
Treatments	I	II	III	IV
1	146.5	177.6	121.5	155.3
2	155.7	240.0	150.0	177.2
3	141.8	158.4	113.8	201.4
4	80.9	241.6	131.9	207.0
EMS	78.17	28.31	108.0	67.9

Analyse the data and draw conclusions.

Analysis:

Step-I: Testing the homogeneity of error variance using Bartlett's

χ^2 - test

$$\bar{s}^2 = \frac{1}{p}\sum_i s_i^2 = 70.60$$

where s^2_i are error mean square based on 12 d.f. each.

$$\chi^2_{(p-1)} = \frac{n[p\log_e \bar{s}^2 - \sum_i \log_e s_i^2]}{1+[(P+1)/3np]}$$

$$\chi^2_3 = \frac{12[4\log_{10}70 \times 2.3026 - 2.3026 \times 7.21012]}{1+[5/(3x12x4)]}$$

$$= \frac{12x0.42616}{1.3047} = 4.94$$

The table value of χ^2 at 5 per cent level of significance for 3 degree of freedom is 7.815. Since calculate value of χ^2 is less than the table value of χ^2, we conclude that error mean squares are not significantly different from each other. Hence we conclude that error variance are homogeneous.

Step-II Combined Analysis of the data

We now make a combined analysis of data treating it at a two-way classification in an ordinary repeated trial.

	Blocks				
Treatments	**I**	**II**	**III**	**IV**	**Total**
1	146.5	177.6	121.5	155.3	600.9
2	155.7	240.0	150.0	177.2	722.9
3	141.8	158.4	113.8	201.4	615.4
4	80.9	241.6	131.9	207.0	661.4
Total	524.9	817.6	517.2	740.9	2600.6

$$C.F. = \frac{(2600.6)^2}{80} = 84539.00$$

$$\text{Total SS} = \frac{451498.86}{5} - 84539.00 = 5760.77$$

SS due to places = 88020.92 - 84539.00 = 3481.92

SS due to treatments = 84991.62 – 84539.00 = 452.62

SS due to treatments interactions

= 5760.77 – 3481.92 – 452.62 = 1826.23

Pooled error MS = 78.23 + 28.31 + 107.94 + 67.90 = 282.38

ANOVA

Source	D.F.	S.S.	M.S.	F
Places	3	3481.92	1160.64	4.11*
Treatment	3	452.62	150.54	
Interaction (Places x treatment)	9	1826.23	220.91	
Pooled error	64		282.38	
Total	79	5760.77		

* Indicate significant at 5% level.

The F ratio for interaction is 220.91/282.38=0.7823.The value is non significant. Assuming interaction to be non significant we can pool the sum of squares for interaction and for error and obtained a more precise estimate of error.

(9x220.91+48x282.38)/(9+48)=272.67. Test the treatment mean square against pooled estimate of experimental error we get F=150.54/272.67=.55 which shows the average treatment differences over the different places is non significant.

Example: With a view to studying the effect of use a spraying on yield of wheat, an experiment was conducted at Agricultural Research Station Arnaj in Gujrat State during 1960-61, 61-62, 62-63 and 63-64. The experimental sites were choosen a fresh every year. The design adopted was randomized block with four replications. The treatments were

N_0 : Control; N_1 : 5.60 Kg/h of N of urea; N_2 : 11.21 Kg/h of N as urea; N_3 : 11.21 Kg/h of N as A/S + 5.60 Kg/h N as urea; N_4: 11.21 Kg/ha of N as A/S + 11.21 Kg/ha of N as urea. Ammonium sulphate was supplied as basal dose and urea applied by spraying in two equal doses. The mean yield in Kg/h is given below:

	Years			
Treatment	1960-61	61-62	62-63	63-64
N_o	863	1442	1248	696
N_1	882	1369	1017	712
N_2	854	1407	1014	756
N_3	860	1363	1199	943
N_4	850	1406	1307	955
EMS	2106.81	43430.56	27989.29	5069.44

Analyse the data and draw conclusions.

Calculation

Testing the homogeneity of error variance using Bartlett's test.

We calculate the following

$$\bar{s}^2 = \frac{1}{p}\sum_i s_i^2 = 19649.025; \qquad \chi_3^2 = 28.443$$

The table value of χ^2 at 5 per cent level of significance for 3 degree of freedom is 7.815. Since calculated value of χ^2 is greater than table value of χ^2. We conclude that error mean squares are significantly different from one another. Hence we can infer that error mean squares are heterogeneous. We shall therefore, analyse the data using weighted method of analysis.

Calculations of various sum of squares.

	Years				
Treatment	1	2	3	4	$\Sigma\ w_i t_{ij}$
N_o	863	1442	1248	696	2.49404
N_1	882	1369	1017	712	2.50387
N_2	854	1407	1014	756	2.48843
N_3	860	1363	1199	943	2.66950
N_4	850	1406	1307	955	2.67890
p_i	4309	6987	5785	4062	
w_i	0.00190	0.00009	0.00014	0.00079	
$w_i p_i$	8.18416	0.62883	0.80990	3.20898	
s_i	3714109	9767779	6765839	3364170	

$$G = \sum_j \left(\sum_i w_i t_{ij}\right) = 12.83474;\ C.F. = \frac{(G.T.)^2}{t\sum_i w_i} = 11282.91$$

Total SS = $\Sigma w_i s_i$ - C.F. = 11540.82 – 11282.91 = 257.91

$$\text{Treatment SS} = \frac{\sum_j \left(\sum_i w_i t_{ij}\right)^2}{\sum_i w_i} - CF$$

= (32.98462/0.00292) – 11282.91 = 13.19

$$\text{SS for years} = \frac{1}{t}\sum_{i=1}^{p} w_i p_i^2 - CF$$

= (57391.9973/5) – 11282.91 = 195.49

SS due to interaction between treatments and year

= Total SS – Treat SS – SS for year

= 257.91 – 13.19 – 195.49 = 49.23

Test for Interaction:

For testing the significance of interaction, transform sum of squares for interaction in to χ^2 using the formula.

χ^2 = [(n-4) (n-2)]/ [n (n+t-3)] (SS due to interaction), with

d.f. = [(p-1) (t-1) (n-4)]/ (n + t – 3)

Thus χ^2 = [8 x 10 x 49.23]/[12 (12 + 5-3)] = 23.44 with 7 d.f. (approx.)

The table value of χ^2 at 5 per cent level of significance for 7 d.f. is 14.067. Since calculated value of χ^2 is greater than table value of χ^2, it can be inferred that interaction is present. Since interaction is present we shall carry out the unweighted analysis.

Unweighted Analysis:

C.F. = 447026449/20 = 22351322.45

Treatment SS = (89609811.00/4) - 22351322.45 = 50905.30

SS for years = (117351719/5) - 22351322.45 = 1119021. 535

Table SS = 1260 574.55

SS due to years x treatment interaction = Table SS – Treat SS – SS for years

= 1260574.55 – 50905.30 – 1119021.35 = 90647.90

Analysis of variance:

ANOVA

Source	D.F.	S.S.	M.S.	F
Treatment	4	50905.30	12726.33	1.68
Years	3	1119021.35	373007.11	49.38**
Treatment x years	12	90647.90	7553.99	
Pooled error	48	78596.10	1637.42	

Since year component comes out to be significant, we can conclude that the season has a significant effect on the trials.

❐ ❐ ❐

Chapter – 13

Balanced Incomplete Block Design

When the number of treatment to be tested is more than the block size, the complete block design are to be abandoned for experimentation and incomplete block designs are needed in such situations. In incomplete block designs, the block size is less than the number of treatments to be tested. These designs were introduced by Yates in order to eliminate heterogeneity to a greater extent than is possible with Randomized Block Design and Latin Square Design when the number of treatment is large. Among incomplete block designs, the Balanced Incomplete Block Designs and Lattice Designs are commonly used by agricultural and biological workers in plant breeding and other trials.

Balanced Incomplete Block Design (BIBD)

A BIBD is an arrangement of v treatment arranged in b blocks each of size k such that.

1. Every treatment occurs atmost once in a block.
2. Every treatment occurs exactly in r blocks.
3. Every pair of treatment occurs together exactly in λ blocks, v, b, r, k, λ are known as the parameter of BIB design and they satisfy the relations.

(i) $vr = bk$ (ii) $\lambda(v-1) = r(k-1)$, (iii) $b \leq v$

The BIB design is said to be symmetrical if $v = b$ and $r = k$.

A plan of BIB design for $v = b = 5$, $r = k = 4$ and $\lambda = 3$ is

Blocks				
(i)	1	2	3	4
(ii)	1	2	3	5
(iii)	1	2	4	5
(iv)	1	3	4	5
(v)	2	3	4	5

The blocks and the treatments with in block, will be randomized in experimental layout. The parameter combinations along with the plans of the BIB designs in the practically useful range are given in the statistical tables by Fisher and Yates (1973), Hafner Pub. Co., New York.

Statistical Analysis

The mathematical model for BIBD (Intra Block Analysis)

$Y_{ij} = \mu + t_i + b_j + e_{ij}$

Where y_{ij} is the observation recorded on the i^{th} treatment in the j^{th} block ($i = 1, 2, .. v, j = 1, 2, \ldots b$)

μ is the general mean

t_i is the effect of the i^{th} treatment

b_j is the effect of the j^{th} block

e_{ij} are the random error components assumed to be distributed normally about mean zero and constant variance σ^2

In the intra block analysis we assume that the treatment effects and the block effects are fixed though unknown and e_{ij}'s are uncorrelated random variables. On minimizing the error sum of squares we get a set of normal equations which can be solved to get the estimate of different contrasts of various estimated treatment and the block effects.

we compute

G = Grand total of observations

$\overline{Y}$ = Grand mean = G/n where n = vr = bk total number of observations.

T_i = Sum of observation for treatment i, (i = 1, 2... v)

B_j = Sum of observation in block j, (j = 1, 2...., b)

C.F.= G^2/n

Q_i = Adjusted total for the i^{th} treatment

= T_i–[Sum of block totals in which treatment i occurs/block size (k)]

The i^{th} effect $\hat{t}_i = (kQ_i)/\lambda v \quad i = 1, 2.... .v.$

Adjusted treatment mean for i^{th} treatment

= i^{th} treatment effect ($\hat{t}_i$) + grand mean $(\overline{Y})$.

$= \frac{Q_i}{rE} + \overline{Y}$ where $E = \lambda v/rk$ and is called the efficiency factor of BIBD.

Various sum of square can be obtained as follows.

(1) Total sum of squares (TSS) = Σ (observation)^2 – C.F.

(2) Treatment Sum of squares unadjusted (SST_u) = $[\Sigma T_i^2/r]$ –C.F.

(3) Block sum of squares unadjusted (SSB_u) = $[\Sigma B_j^2/k]$ –C.F.

(4) Treatment sum of square adjusted (SST_a)

$$= \Sigma \hat{t}_i Q_i = \sum_i \frac{k}{\lambda v} Q_i^2 \quad = \Sigma (kQ_i)^2/\lambda vk$$

The treatment sum of square can be calculated as

Treatment Number	Treatment total (T_i)	Blocks where i^{th} treatment occurs	Sum of the blocks containing i^{th} treatment $\sum_{j(i)} B_j$	$kQ_i = kT_i - \sum_{j(i)} B_j$	$(kQ_i)^2$
1 2 .. v					
Total					$\Sigma(kQ_i)^2$

(5) Error SS (SSE) = Total SS–Block SS (unadjusted)-Treat SS (adjusted)

ANOVA for a BIBD (v, b, r, k, λ)

Source	DF	SS	MS	F
Block (unadjusted)	b-1	SSB_u		
Treatment (adjusted)	v-1	SST_a	MST	$\frac{MST}{MSE} \sim F(v-1, bk-b-v+1)$
Error (Intra block)	bk-b-v+1	By subtraction SSE	MSE	
Total	bk-1	TSS		

Here $MST = SST_a/v-1$ and $MSE = SSE/(bk - b-v+1)$

As usual if $F_{cal} < F_{tab}$, then treatments do not differ significantly. But if $F_{cal} \geq F_{tab}$ then treatments differ significantly and we can proceed further to find the C.D. value.

$$\text{Coefficient of variation} = \sqrt{\frac{MSE}{\bar{Y}}} \times 100$$

$$\text{S.E. of different between two adjusted treatment means} = \sqrt{\frac{2MSE}{rE}}$$

$$CD \text{ at } 5\ \% = t_{(a,\ \text{error d.f.})} \times \sqrt{\frac{2MSE}{rE}}$$

In case one is interested to make inferences about block effects, the adjusted block mean square is to be compared with error mean square. The adjusted block SS can be computed using the following identify.

SS due to blocks adjusted (SSB_a) + SS due to treatments unadjusted (SST_u)

= SS due to Blocks unadjusted (SSB_u) + SS due treatment adjusted (SST_a)

Here SS due to treatment unadjusted (SST_u) = $\Sigma T_i^2 / r - C.F.$

ANOVA

Source	DF	SS	MS	F
Block (adjusted)	b-1	SSB_a		
Treatment (adjusted)	v-1	SST_u	E_b	E_b/E_e
Error (Intra block)	bk-b-v+1	SSE	E_e	
Total	bk-1			

If $F_{cal} > F$ ([α, (b-1), (bk-v-b+1)] then the block effects are significant.

In BIBD each pair of treatment means are estimated with the same precision.

Efficiency

The variance of difference between any two estimated treatment effects is given by $V(\hat{t}_i - \hat{t}_m) = \frac{2\hat{\sigma}_1^2}{rE}$

If a randomized block design with v treatments and r replications has been adopted, the corresponding variance would have been

$$V(\hat{t}_i - \hat{t}_m) = \frac{2\hat{\sigma}_1^{\,2}}{r}$$

where $\hat{\sigma}_1^{\,2}$ is the per observation variance in case of RBD. Thus the efficiency of BIBD as compared to RBD is given by

$$\text{Efficiency} = \frac{2\hat{\sigma}_1^{\,2}}{r} \Big/ \frac{2\hat{\sigma}^2}{rE} = E\frac{\hat{\sigma}_1^{\,2}}{\hat{\sigma}^2}$$

The quantity E is called efficiency factor of a BIB design

Here E<1 since k<v.

SE of difference between the adjusted treatment means is

$$S.E_d = \sqrt{\frac{2E_e}{rE}}$$

Critical Difference (C.D.) = $S.E_d \times t_{(\alpha,\ \text{error d.f.})}$

Analysis with recovery of inter block information

In the intra block analysis it is assumed that the treatment and block effects are fixed. The treatment effect may be treated as fixed but the block effect should be treated as random variable because the block effect in an incomplete block design depend upon the set of treatments received by the block. Therefore, in BIBD, between block variation on the block comparisons provide error variance.

Also the block comparisons contains some information on treatment comparisons (as some set of treatment occurs in different blocks) which can be extracted, if we can estimate the variance between blocks with in a replicate which are treated alike.

If the Inter block information is recovered, the adjusted treatment mean is equal to the unadjusted treatment mean plus a correction.

If $E_b > E_e$ the adjusted totals can be obtained a $T_i^* = T_i + \mu w_i$

$$\text{Where } \mu = \frac{(b-1)(E_b - E_e)}{v(k-1)(b-1)E_b + (v-k)(b-v)E_e}$$

$$\text{and } w_i = (v-k)T_i - (v-1)\sum_{j(i)} B_j + (k-1)G$$

Then S.S. due to treatment adjusted

$$= \frac{\Sigma T_i^{*2}}{r} - \frac{(\Sigma T_i^*)^2}{vr}$$

and MS due to treatment (adjusted)

$$= \frac{1}{v-1}\left[\frac{\Sigma T_i^{i2}}{r} - \frac{(\Sigma T_i^*)^2}{vr}\right]$$

When inter block information is recovered an approximate F-test is

$$F\,[(v-1),\ bk-v-b+1)\ \text{d.f.}\ ]$$

$$= \frac{\text{MS among adjusted means}}{\text{average effecitve error variacne}}$$

Effective error variance = E_e [1+ (v-k)μ]

Example

An experiment on 13 varieties in paddy was conducted at Agricultural Research Institute, Rajendra Nagar, Andhra Pradesh, to find out the best variety suitable for the tract. The experiment was laid out in BIB design with 13 blocks of 4 plots each and each treatment replicated four times. The Yield data is given as:

	Yield of paddy (in kg/ha)			
Blocks				
I	23.6	18.9	28.0	24.3
II	19.5	32.3	22.0	21.8
III	30.7	725.3	20.3	17.2
IV	25.3	26.0	36.6	18.4
V	20.4	32.5	31.8	33.4
VI	32.6	31.4	28.2	32.5
VII	33.7	31.2	26.7	31.6
VIII	33.5	32.5	32.3	35.9
IX	37.2	31.9	38.4	30.3
X	27.7	29.7	25.9	34.5
XI	36.0	30.1	31.1	29.4
XII	32.8	32.7	28.8	40.7
XIII	31.3	30.5	39.5	27.7

Analyse the data

(i) without recovery of intra block information.

(ii) with recovery of inter block information.

Calculations:

The parameters of the given design are

b = 13 = number of incomplete blocks.

v = 13 = number of varieties of paddy.

k = 4 = number of plots per block.

r = 4 = number of times a variety is replicated.

λ = 1 = number of blocks in which a pair of treatment occurs together in the given design.

(i) Calculation of Q_i:

Sr. No.	Treatment total T_i	Block Total B_j	Serial No. of block where i^{th} treatment occurs	Sum of the block which contain i^{th} treatment $\sum_{j(i)} B_j$	kT_i	KQ_i= (6)-(5)	$(kQ_i)^2$
1	2	3	4	5	6	7	8
1	139.7	94.8	7,9,11,12	521.6	558.8	37.2	1383.8
2	116.2	65.6	4,9,10,13	490.9	464.8	-26.1	681.2
3	118.2	93.5	1,2,8,9	462.4	472.8	10.4	108.2
4	111.9	106.3	2,6,10,11	463.7	447.6	-16.1	259.2
5	121.1	118.1	4,6,7,8	488.4	484.4	-4.0	16.0
6	110.3	124.7	1,6,12,13	483.5	441.2	-42.3	1789.3
7	122.3	123.2	5,8,11,13	506.9	489.2	-17.7	313.3
8	130.2	134.2	2,4,5,12	455.0	520.8	65.8	4329.6
9	112.4	137.8	1,5,7,10	453.9	449.6	-4.3	18.5
10	113.4	117.8	3,5,6,9	474.1	453.6	-20.5	420.3
11	91.7	125.6	1,3,4,11	420.2	366.8	-53.4	2851.6
12	111.8	135.0	2,3,7,13	441.3	447.2	5.9	34.8
13	136.4	129.0	3,8,10,12	480.5	545.6	65.1	4238.0
	1535.6	1535.6		6142.4		0.0	16443.8

$$C.F. = \frac{\left(\sum_i \sum_j Y_{ij}\right)^2}{bk} = \frac{(1535.6)^2}{13 \times 4} = 45347.4$$

$$\text{Total SS} = \sum_i \sum_j Y_{ij}^2 - C.F.$$

$$= 46932.20 - 45347.40 = 1584.8$$

$$\text{Block SS (unadjusted)} = \sum_j \frac{B_j^2}{k} - \text{C.F.}$$

$$= 46090.2 - 45347.4 = 742.8$$

$$\text{Treatment SS (adjusted)} = \frac{\sum_i (kQ_i)^2}{vk\lambda}$$

$$= \frac{16443.8}{4x1x13} = 316.2$$

$$\text{Treatment SS (unadjusted)} = \sum_i \frac{T_i^2}{r} - \text{C.F.}$$

$$= 45814.6 - 45347.4 = 467.2$$

Error SS (Intra Block) = Total SS – Treatment SS (adjusted)
– Block SS (unadjusted)

$$= 1584.8 - 316.2 - 742.8 = 525.8$$

(ii) Analysis of Variance

ANOVA

Source	D.F.	S.S	M.S.	F
Block (unadjusted)	12	742.80	61.9	1.4
Treatment (adjusted)	12	316.20	26.40	0.6
Error	27	525.80	43.8	-
Total	51	1584.80		

Here the treatments do not differ significantly

Standard error of the difference between two variety means (adjusted)

$$\text{Efficiency factor of BIBD, } E = \frac{\lambda v}{rk} = \frac{13}{16} = 0.8125$$

Analysis with recovery of inter block information

i) Block SS (adjusted)	=	Block SS (unadjusted) + Treatment SS (adjusted) – Treatment SS (unadjusted).
	=	742.8 + 316.2 – 525.8
	=	533.3

(ii) Analysis of variance

Source	D.F.	S.S.	M.S.	F
Block (adjusted)	12	533.30	44.40 E_b	1.00
Treatment (unadjusted)	12	467.20	38.90	0.90
Error	27	525.80	43.80 E_e	-
Total	51	1584.80		

We have $E_b = 44.4$ and $E_e = 43.8$

Here $E_b > E_e$ we can proceed for the analysis of inter block information

(iii) Calculation of μ and w_i

$$\mu = \frac{(b-1)(E_b - E_e)}{v(k-1)(b-1)E_b + (v-k)(b-v)E_e} = \frac{12 \times 0.60}{39 \times 12 \times 44.40} = 0.0003$$

$$w_i = (v-k)\,T_i - (v-1)\sum_{j(i)} B_j + (k-1)G$$

Calculation of w_i using above equation

Treatment No.	Treatment total T_i	Serial No. of block where i^{th} treatment occurs	Sum of the block totals where i^{th} treatment occurs	w_i	$T_i^* = T_i + \Box w_i$	Adjusted mean (6)/r
1	2	3	4	5	6	7
1.	139.7	7,9,11,12	521.6	841.5	139.4	34.9
2.	116.2	4,9,10,13	490.9	75.0	116.2	29.1
3.	118.2	1,2,8,9	462.4	-3.0	118.2	29.5
4.	111.9	2,6,10,11	463.7	242.7	112.0	28.0
5.	121.1	4,6,7,8	488.4	-116.1	121.1	30.3
6.	110.3	1,6,12,13	483.5	305.1	110.4	27.6
7.	122.3	5,8,11,13	506.9	162.9	122.3	30.6
8.	130.2	2,4,5,12	455.0	-471.0	130.1	32.5
9.	112.4	1,5,7,10	453.9	223.2	112.5	28.1
10.	113.4	3,5,6,9	474.1	184.2	113.5	28.4
11.	91.7	1,3,4,11	420.2	1030.5	92.0	23.0
12.	111.8	2,3,7,13	441.3	246.6	111.9	28.0
13.	136.4	3,8,10,12	480.5	-712.8	136.2	34.0
				Total	1535.6	

$$\text{Treatment SS (adjusted)} = \sum_i \frac{T_i^{*2}}{r} - \left(\sum_i T_i^*\right)^2 \Big/ vr$$

$$= 45803.70 - 45347.40 = 456.30$$

M.S. due to treatment (adjusted)

$$= \frac{456.30}{12} = 38.02$$

$$F\ (12 \text{ and } 27 \text{ d.f.}) = \frac{\text{M.S.(adjusted) due to treatment}}{\text{Average effective error variance}}$$

$$= \frac{38.02}{43.93} = 0.86\ (\text{NS})$$

where average effective error variance = $E_e\ [1+ (v-k)\mu]$

$$= 43.80 \times 1.003 = 43.93$$

The treatment effects are not significantly different among themselves.

❐ ❐ ❐

Chapter – 14

Lattice Designs

Balanced incomplete block (BIB) designs are the most efficient designs in the class of binary incomplete block designs but these designs require usually a large number of replications and are not available for all combinations of parameters values. To over come these difficulties Yates (1936) introduced a class of designs known as Lattice Designs. These designs are most commonly used in agricultural research. There is sufficient flexibility in the design to make its application simpler than most other incomplete block designs. The characteristic features of these designs are that the number of treatment is a perfect square and the block size is the square root of this number. We have balanced lattice and the partially balanced lattice designs. Both require that the number of treatments must be a perfect square.

Balanced lattice

The balanced lattice design is characterized by the following basic features.

1. The number of treatment (v) must be a perfect square (i.e. $v = s^2$, such as 25, 36, 49, 64, 81, 100 etc.). As the number of treatment becomes large, adding a two more or eliminating some less important treatments is usually easy to accomplish. For example if a plant breeder wishes to test the performance of 80 varieties in a balanced lattice design, all he needs to do is add one more

variety for a perfect square or has 82 or 83 varieties to start he can easily eliminate one or two.

2. The block size (k) is equal to the square root of the number of treatment $\left(\text{i.e. } k = \sqrt{v}\right)$.

3. The number of replications (r) is one more than the block size [i.e. r = (k+1)]. The number of replications required is 6 for 25 treatment, 7 for 36 treatment and so on.

Randomization and layout

We illustrate the randomization and layout of a balanced lattice design with a field experiment involving nine treatments. There are four replications, each consisting of three incomplete blocks with each block containing three experimental plots.

Step-I Divide the experimental area in to r = (k+1) replications, each containing $v = k^2$ experimental plot.

	Replication I			Rep. II			Rep. III			Rep. IV		
Block 1	1	2	3	10	11	12	19	20	21	28	29	30
Block 2	4	5	6	13	14	15	22	23	24	31	32	33
Block 3	7	8	9	16	17	18	25	26	27	34	35	36

Step-II Divided each replication in to k incomplete blocks each containing k experimental plots.

Step-III Select a basic balanced lattice plan corresponding to the number of treatments to be tested. For our example, the basic plan to 3x3 balanced lattice design is

Incomplete Block Number	Treatments Numbers														
	Rep. I				Rep. II				Rep. III				Rep. IV		
1	1	2	3		1	4	7		1	5	9		1	6	8
2	4	5	6		2	5	8		2	6	7		2	4	9
3	7	8	9		3	6	9		3	4	8		3	5	7

Step IV Randomize the replication arrangement of the selected basic plan following an appropriate randomization scheme. The

outcome of new plan at this step is

Incomplete Block Number	Treatments Numbers															
	Rep. I				Rep. II				Rep. III				Rep. IV			
1	1	4	7		1	2	3		1	5	9		1	6	8	
2	2	5	8		4	5	6		2	6	7		2	4	6	
3	3	6	9		7	8	9		3	4	8		3	5	7	

Step V Randomize the incomplete blocks with in each replication. After four independent randomization processes, the reassigned incomplete blocks may be shown a.

Incomplete Block Number in Basic Plan	Reassigned Incomplete Block Number in New Plan															
	Rep. I				Rep. II				Rep. III				Rep. IV			
1		3				2				3				1		
2		2				1				1				3		
3		1				3				2				2		

The out come of the new plan at this step is.

Incomplete Block Number	Treatments Numbers															
	Rep. I				Rep. II				Rep. III				Rep. IV			
1	3	6	9		4	5	6		3	5	7		1	5	9	
2	2	5	8		1	2	3		1	6	8		3	4	8	
3	1	4	7		7	8	9		2	4	9		2	6	7	

Step-VI: Randomize the treatment arrangement with in each incomplete block, the reassigned treatment sequence may be shown as:

Incomplete Block Number	Treatments Numbers															
	Rep. I				Rep. II				Rep. III				Rep. IV			
1	9	3	6		5	4	6		7	5	3		1	9	5	
2	8	5	2		3	2	1		6	8	1		4	8	3	
3	7	1	4		9	8	7		2	4	9		7	2	6	

Step-VII: Apply the final outcome of the randomization process of step-6 to field layout. The resultant layout of a 3x3 balanced lattice designs involving nine treatments T_1, T_2,..... T_9 becomes.

	Replication I			Rep. II			Rep. III			Rep. IV		
Block 1	T_9	T_3	T_6	T_5	T_4	T_6	T_7	T_5	T_3	T_1	T_9	T_5
Block 2	T_8	T_5	T_2	T_3	T_2	T_1	T_6	T_8	T_1	T_4	T_8	T_3
Block 3	T_7	T_1	T_4	T_9	T_8	T_7	T_2	T_4	T_9	T_7	T_2	T_6

Here important feature in the layout of balanced lattice design is that every pair of treatment occurs.together only once in the same block. For example, treatment 1 appears only once with treatment 4 and 7 in block 3 of replication I, with treatment 2 and 3 in block 2 of replication in II with treatment 6 and 8 in block 2 of replication III and with treatment 5 and 9 in block I of replication IV. As a consequence of this feature, degree of precision for comparing each pair of treatment in a balanced lattice design is the same for all the pairs.

Analysis: There are four sources of variation that can be accounted for in a balanced design: replication, treatment, incomplete block and experimental error. As compared to RBD, the incomplete blocks is an additional source of variation and reflects the differences among incomplete blocks of the same replication.

The data of the 16 rice fertilizer treatments in 4 x4 balanced lattice designs with five replications given below has been used for computational procedure for the analysis of variance.

Tillers number per square meter from 16 fertilizer treatments tested in a 4 x 4 Balanced Lattice Design.

Blocks Numbers	Tillers No/m^2				Block Totals
	Rep I				
1	(1) 147	(2) 152	(3) 167	(4) 150	616
2	(5) 127	(6) 155	(7) 162	(8) 172	616
3	(9) 147	(10) 100	(11) 192	(12) 177	616
4	(13) 155	(14) 195	(15) 192	(16) 205	747
Rep. Total R_1					2595

Block Numbers (B)	Tillers No/m^2				Block Totals (B)
	Rep. II				
5	(1) 140	(5) 165	(9) 182	(13) 152	639
6	(10) 97	(2) 155	(14) 192	(6) 142	586
7	(7) 155	(15) 182	(3) 192	(11) 192	721
8	(16) 182	(8) 207	(12) 232	(4) 162	783
Rep. Total R_2					2729

Blocks Numbers	Tillers No/m^2				Block Totals
	Rep III				
9	(1) 155	(6) 162	(11) 177	(16) 152	646
10	(5) 182	(2) 130	(15) 177	(12) 165	654
11	(9) 137	(14) 185	(3) 152	(8) 152	626
12	(13) 185	(10) 122	(7) 182	(4) 192	681
Rep. Total R_3					2607

Blocks Numbers	Tillers No/m^2				Block Totals
	Rep. IV				
13	(1) 220	(14) 202	(7) 175	(12) 205	802
14	(13) 205	(2) 152	(11) 180	(8) 187	724
15	(5) 165	(10) 150	(3) 200	(16) 160	675
16	(9) 155	(6) 177	(15) 185	(4) 172	689
Rep. Total R_4					2890

Blocks Numbers	Tillers No/m^2				Block Totals
	Rep V				
17	(1) 147	(10) 112	(15) 177	(8) 147	583
18	(9) 180	(2) 205	(7) 190	(16) 167	742
19	(13) 172	(6) 212	(3) 197	(12) 192	773
20	(3) 177	(14) 220	(11) 205	(4) 225	827
Rep. Total R_5					2925

The values in parenthesis correspond to the treatment numbers.

Step-1: Calculate the block totals (B) and replication totals (R) treatment totals (T) and the grand Total (G).

Step-2: For each treatment calculate the B_t value as the sum of block totals over all blocks in which the particular treatment appears. For example treatment 5 occurs in blocks 2, 5, 10, 15, and 20. Thus B_t for treatment 5 is computed is the sum of the blocks totals of the blocks 2, 5, 10, 15 and 20 or $B_5 = 616 + 639 + 654 + 675 + 827 = 3411$.

Note that sum of all B_t values over all the treatments must be equal to (K) (G) where K is the block size.

Table: Computations of the adjusted and unadjusted treatment totals for the 4 x 4 balanced lattice design.

Treatment Number	Treatment Total (T)	Block Totals (B_t)	W = 4T-5 B_t + G	T′ = T+ μW	M′ = T′/5
1	809	3286	552	829	166
2	794	3322	312	805	161
3	908	3411	323	920	184
4	901	3596	-630	878	176
5	816	3411	-45	814	163

Treatment Number	Treatment Total (T)	Block Totals (B_t)	W = 4T-5 B_t + G	T′ = T+ μW	M′ = T′/5
6	848	3310	588	869	174
7	864	3562	-608	842	168
8	865	3332	546	885	177
9	801	3312	390	815	163
10	581	3141	365	594	119
11	946	3534	-140	941	188
12	971	3628	-510	953	191
13	869	3564	-598	848	170
14	994	3588	-218	986	197
15	913	3394	428	928	186
16	866	3593	-755	839	168
Sum	13746(G)	54984	0	-	-

Step-3: For each treatment, calculate

$W = KT - (k+1)\, B_t + G$

For example the value of W for treatment 5 is computed as

$W_5 = 4\,(816) - (5)\,(3411) + 13746 = -45$ sum of W values over all the treatments must be zero.

Step-4: Compute the total SS, the replication SS and the treatment (unadjusted) SS as.

$$C.F. = \frac{G^2}{k^2(k+1)} = \frac{(13{,}746)^2}{(16)(5)} = 2361906$$

$$\text{Total SS} = \sum X^2 - CF$$

$$= [(147)^2 + (152)^2 \ldots\ldots\ldots\ldots + (225)^2] - 2\,361906 = 58856$$

$$\text{Replication SS} = \frac{\sum R^2}{k^2} - CF$$

$$= \left[\frac{(2595)^2 + (2729)^2 \ldots\ldots + (2925)^2}{16}\right] - 2361906 = 5946$$

$$\text{Treatment (unadj) SS} = \frac{\sum T^2}{(k+1)} - CF$$

$$= \left[\frac{(809)^2 + (794)^2 ..... + (866)^2}{5}\right] - 2361906 = 26995$$

Step-5: Compute the block (adjusted) SS [i.e., the sum of square for blocks with in replication adjusted for treatment effects) as:

$$\text{Block (adj) SS} = \frac{\sum_{i=1}^{v} w_i^2}{(k^3)(k+1)}$$

$$= \left[\frac{(552)^2 + (312)^2 ..... + (-755)^2}{64 \times 5}\right] = 11382$$

Step-6: Intra block error SS=Total SS–Rep SS-Treat (unadj) SS-Block (adj) SS

$= 58856 - 5946 - 26995 - 11392 = 14533$

Step-7: Complete the block (adj) mean square and the intra block error mean square as:

$$\text{Block (adj) MS} = \frac{\text{Block (adj) SS}}{k^2 - 1} = \frac{11382}{15} = 759$$

$$\text{Intra block error MS} = \frac{\text{Intra block error SS}}{(k-1)(k^2-1)} = \frac{14533}{(3)\ (15)} = 323$$

Step-8: For each treatment calculate the adjusted treatment totals T' as

$$T' = T + \mu W$$

$$\text{where } \mu = \frac{\text{Block (adj) MS} - \text{Intra block error MS}}{k^2\ [\text{Block (adj) MS}]}$$

If the intra block error MS is greater than the block (adj) MS, μ is taken to be zero and no adjustment for treatment is necessary. The F test for significance of treatment effect is then made in the usual manner as the ratio of treatment(unadj) MS and intra block error MS and steps 8 to 12 and step 15 can be ignored.

In our example the intra block error MS is smaller than the block (adj) MS. The adjustment factor μ is computed as $\mu = \frac{759 - 329}{16\,(759)} = .0359$. The T′ value for treatment 5 is computed as

$T_5' = 816 + (.0359)\,(-45) = 814.$

Step-9: For each treatment, calculate the adjusted treatment mean M′ as

$$M' = \frac{T'}{k+1} \quad \text{Here } M_5' = \frac{814}{5} = 163$$

Step 10: Compute the adjustment treatment mean square as

$$\text{Treatment (adj) MS} = \left[\frac{1}{(k+1)\,(k^2-1)}\right]\left[\sum T'^2 - \frac{G^2}{k^2}\right]$$

$$= \left[\frac{1}{(5)\,(15)}\right]\left\{\left[(829)^2 + (805)^2 + \ldots\ldots + (839)^2\right] - \frac{(13746)^2}{16}\right\}$$

$= 1602$

Step 11: Complete the effective error MS as

Effective error MS = (Intra Block error MS) $(1 + k\mu)$

$= 323\,[1+4\,(0.\,0359)] = 369$

$$\text{C.V.} = \sqrt{\frac{\text{Effective error M.S.}}{\text{grand mean}}} \times 100 = \sqrt{\frac{369}{172}} \times 100 = 11.2\%$$

Step 12: Compute the F value for testing the treatment differences as:

$$F = \frac{\text{treatment (adj) MS}}{\text{Effective Error MS}} = \frac{1602}{369} = 4.34$$

Step 13: Compare the computed F value to the tabular F values with $n_1 = (k^2\text{-}1) = 15$ and $n_2 = (k\text{-}1)\,(k^2\text{-}1) = 45$ degree of freedom. As the computed F value is larger then the tabular F values (2.47) at 1% level of significance, the treatment difference is judged to be highly significant.

Step 14: Construct the analysis of variance table

Source of variation	Degree of freedom	Sum of squares	Mean squares	Comp Comp uted F
Replication	k = 4	5946		
Treatment (unadj)	k^2-1=15	26995		
Block (adj)	k^2-1=15	11382	759	
Intra block error	(k -1) (k^2-1) = 45	14533	323	
Treatment (adj)	(k^2-1) = 15	-	1602	4.34**
Effective error	(k-1) (k^2-1)=45	-	369	
Total	k^2(k+1) -1 =79	58856		

Step-15: Estimate gain in precision of a balanced lattice design relative to RBD design as:

$$R\,E = \frac{100\,[\text{Block (adj) SS} + \text{Intra block error SS}]}{k\,(k^2 - 1)\,(\text{Effective error MS})}$$

$$= \frac{100\,(11,382 + 14,533}{4\,(16 - 1)\,(369)} = 117\%$$

Thus the use of 4 x 4 balanced lattice design is estimated to have increased the experimental precision by 17% over that which would have been obtained with RBD.

Partially Balanced Lattice

The partially balanced lattice design is similar to the balanced lattice design but allows a more flexible choice of number of replications. Here also the treatment number must be a perfect square and the block size is equal to the square root of this treatment number, the number of replication is not prescribed as a function of number of treatment. In fact any number of replication can be used in a partially balanced design.

With two replications, the partially balanced lattice design is

referred as simple lattice, with three replications a triple lattice and with four replications a quadruple lattice and so on. Such flexibility in the choice of number of replications results in a loss of symmetry in the arrangement of treatment over blocks (i.e. some treatment pairs never appear together in the same incomplete block). Consequently, the treatment pairs that are tested in the same incomplete block are compared with a level of precision that is higher than those that are not tested in the same incomplete block. So there is more than one level of precision for comparing treatment means and data analysis become more complicated.

❐ ❐ ❐

Chapter – 15

Layout and Analysis of Augmented Designs

In Genetic resources environment, an essential activity is to test or evaluate the new germplasm/provenances/superior selections (new treatments with the existing provenances or released varieties (checks or control treatments). A problem in these evaluation studies is that quantity of the genetic material collected from the exploration trips is very limited or can not be made available since a part of this is to be deposited in Genebank. The available quantity of seed is often not sufficient for replication. Moreover the number of new germplasm or provenances to be tested is very high (usually about 1000-2000 and some time up to 3000 associations) and it is very difficult to maintain block homogeneity. RBD can not be used as soil hetrogenity becomes unmanageable. In absence of error term (error variance) in the ANOVA, without replication, test of significance cannot be applied. Both these problems can be surmounted by employing Augmented (or Honnuiaku) designs. Because of simplicity and usefulness, these designs have applications in the fields of plant breeding, entomology, pathology, chemistry, physiology, agronomy and perhaps others for screening experiments on new material and preliminary testing of experiments on promising material. Thus we have to design an experiment in which the experimental material for new (test) treatments is enough only for single replication. However, to ensure connectness property of the design i.e. estimation of all possible paired comparisons and estimation of error, at least two

replications on all treatments are required in traditional block designs viz. randomized complete block design, balanced incomplete block designs, etc. In augmented designs, connectness is ensured by augmenting any standard connected design in control treatments with new (test) treatments and replication of the control provide the estimate of error. More precisely, an augmented experimental design is any standard experimental design in controls augmented with additional treatments in the complete block, incomplete block, the row, the column etc. The blocks in an augmented design may be of unequal size. Federer (1956) proposed three models.

1. Augmented completely randomized design (Augmented Design I).
2. Augmented randomized complete block design (Augmented Design II).
3. Augmented Latin Square Design (augmented Design III).

Only the first two design (I and II) satisfying the above two situations are widely used by plant breeders, hence elaborated below:

Augmented Design I

Concept: When number of seeds is a limitation, the augmented completely randomized design (I) is most suited. In this design the whole experimental area is divided in to N plots (where N is equal to the number of test genotypes (V) + number of checks (C) which are standard varieties or hybrids of known performance) repeated b times, i.e. N = V + bC; and total number of entries, e = V+C. All the N plots are allotted randomly for all V and C, the later repeated b times. Blocks are not marked at all. A representative field design with V=8 (V_1 to V_8). C = 4 (C_1 to C_4), b = 3, e = 12 and N = 20 in a linseed trial is depicted below:

Table 1: Field layout and data collection: Augmented design I

C_3	V_6	C_4	C3	V1	C4	V7	C4	V4	C2
C_1	V_3	C2	C1	C3	V5	C1	V8	C2	V2

The following data on number of tillers per plant were collected on 5 random plants per plot and the average over sample is presented in table 2.

Table 2: Average data on number of tillers in linseed varieties and checks under augmented design I

Checks (Ci)	Symbolic data					Actual data				
	Repeats (b_j)			Total	Mean	Repeats			Total	Mean
	1	2	3	(T_{ci})	($\bar{C}_i$)	1	2	3		
C_1	x_1	x_2	x_3	Tc_1	$\bar{C}_1$	10	8	6	24	8.0
C_2	x_4	x_5	x_6	Tc_2	$\bar{C}_2$	3	5	6	14	4.7
C_3	x_7	x_8	x_9	Tc_3	$\bar{C}_3$	10	9	8	27	9.0
C_4	x_{10}	x_{11}	x_{12}	Tc_4	$\bar{C}_4$	15	14	18	47	15.7
Total	(C)			Tc	C				112	9.33

Varieties (Vi)	Symbolic data (Vi)	Actual data
V_1	x_{13}	6
V_2	x_{14}	40
V_3	x_{15}	4
V_4	x_{16}	24
V_5	x_{17}	11
V_6	x_{18}	18
V_7	x_{19}	8
V_8	x_{20}	30
Total (V)	Tv	141

Grand Total GT = (V+bc) = 141+112 = 253 $\bar{V}$ = T/V = 141/8 = 17.62

Table 3 : The skeleton of ANOVA for augmented design I.

Source of variation	d.f.	SS	MSS
Among entries (e)	e-1 = 11	eSS	eMS
Among checks (C)	C-1 = 3	cSS	cMS
Among varieties (V)	V-1 = 7	vSS	vMS
C Vs. V	1	cvSS	cvMS
Error	C (b-1) = 8	ESS	EMS
Total	N-1 = 19	TSS	

Statistical analysis: All the components of ANOVA can now be developed as under:

i) GCF (General correction factor) = $(GT)^2/N$

$= (253)^2/20 = 3200.4$

ii) $TSS = \sum_{i=1}^{c}\sum_{j=1}^{b} C_{ij}^2 + \sum_{i=1}^{v} Vi^2 - GCF = x_1^2 + x_2^2 + -- + x^2_{12} +$

$x_{13}^2 + --- + x_{20}^2 - GCF$

$= 10^2 + 8^2 + --- + 14^2 + 18^2 + 6^2 + 40^2 + --- + 30^2 - 3200.4$

$= 4897.0 - 3200.4 = 1696.6$

iii) $eSS = \sum_{i=1}^{c} T_{Ci}^2/3 + \sum_{i=1}^{v} V_i^2 - GCF$

$= (Tc_1^2 + Tc_2^2 + --- + Tc_4^2) / 3 + Vi^2 + --- + V_8^2 - GCF$

$$= \frac{24^2+14^2+27^2+47^2}{3} + 6^2 + 40^2 + --- + 30^2 - 3200.4$$

$= 1236.6 + 3637.0 - 3200.4 = 1673.2$

$eMS = eSS/(e-1) = 1673.2/11 = 152.1$

iv) $cCF = T_C^2 / bc = 112^2 / (3 \times 4) = 112^2/12 = 1045.3$

v) $cSS = \sum_{1}^{c} T_{Ci}^2/3 - cCF$

$$= \frac{24^2+14^2+27^2+47^2}{3} - 1045.3 = 1236.6 - 1045.3 = 191.3$$

vi) $vCF = Tv^2/v = 141^2/8 = 2485.1$

vii) $vSS = \sum_{1}^{v} Vi^2 - vCF = 6^2 + 40^2 + --- + 8^2 + 30^2 - 2485.1$

$= 3637.0 - 2485.1 = 1151.9$

$vMS = vSS / (v-1) = 1151.9 / 7 = 164.5$

viii) cvSS = eSS – cSS – vSS = 1673.2 – 191.3 – 1151.9 = 330.0

cvMS = cvSS/1 = 330.0/1 = 330.0

ix) ESS = TSS – eSS = 1696.6 – 1673.2 = 23.4

EMS = ESS / c (b-1) = 23.4/4 (3-1) = 23.4/8 = 2.9

x) F (calculated) = eMS / EMS (for entries) = 152.1/2.9 = 52.4

= cMS / EMS (for checks) = 63.7/2.9 = 21.9

= VMS / EMS (for varieties) = 164.5 / 2.9 = 56.7

= cvMS / EMS (for checks Vs Var S) = 330.1 / 2.9 = 113.8

xi) Standard error of mean difference :

SEd_1 (between 2 checks) =

$$\sqrt{\frac{2EMS}{b}} = \sqrt{\frac{2\times 2.9}{3}} = \frac{2.4}{1.73} = 1.38$$

SEd_2 (between 2 vars.) = $\sqrt{2EMS} = \sqrt{2\times 2.9} = 2.40$

SEd_3 (between a check and variety) =

$$\sqrt{EMS\left(1+\frac{1}{b}\right)} = \sqrt{2.9\left(1+\frac{1}{3}\right)}$$

$$= \sqrt{2.9\times 1.33} = \sqrt{3.86} = 1.96$$

CD = corresponding SEd x t (5% or 1%) for error d.f. (=8)

Presentation of results :The whole ANOVA is presented now as under Table 4.

Table 4 : ANOVA (Augmented Design I)

Source of variation	d.f.	SS	MSS	F
Among entries	11	1673.2	152.1	52.4**
Among checks	3	191.3	63.7	21.9**
Among varieties	7	1151.9	164.5	56.7**
Checks Vs. varieties	1	330.0	330.0	113.8**
Error	8	23.4	2.9	---
Total	19	1696.6		

** $p<0.01$ (significant at 1% level of significance)

With this table of ANOVA, table of means must be presented as

shown below (Table 5)

Table 5 : Mean Performance.

Checks / Var.	No. of tillers per plant	Vars.	No. of tiller / plant
C_1	8.0	V_5	11
C_2	4.7	V_6	18
C_3	9.0	V_7	8
C_4	15.7	V_8.	30
$\bar{C}$	9.33	$\bar{V}$	17.62
V_1	6		
V_2	40	CD_1 at 5%	1.38 x 2.306 = 3.182
V_3	4	CD_2 at 5%	2.40 x 2.306 = 5.5.34
V_4	24	CD_3 at 5%	1.96 x 2.306 = 4.519

Interpretation of results

(i) All the variance among entries / checks / varieties and the single degree of freedom comparison (checks Vs. varieties) are highly significant ($p<0.01$)

(ii) Test variety V_2 and V_8 are significantly highest tillering and V_3 and V_1 the lowest.

(iii) Test varieties are significantly superior over checks ($\bar{C} = 9.33 < \bar{V} = 17.62$).

Augmented design I can be extended to any number of V and C on the same line as above.

Augmented Design II

Concept: This is a slightly modified version of augmented design

I, where instead of repeating all the checks b times throughout the field, the whole field is divided distinctly into b-blocks. Then randomization is done such that all the checks and a part of test genotypes fall only once in each block. Thus the total number of plots (N) remains, the same as in design I (if same number of V and C are used) but the number of plots in each block is variable. The advantage of such a stratification (block formation) of area is to reduce the EMS by apportioning the block effect which is confounded with error in design I. Actually, what RBD does by replicating only the checks (C), thus saving space.

Field plan and randomization: Thus can be illustrated using the example given in augmented design I i.e. V = 8 (V_1 to V_8), C = 4 (C_1 to C_4) b = 3, e = 12 and N = 20. Now divide the whole experimental field into (b=)3 distinct blocks.

Block 1, consisting of (say) 7 plots (C_1 to C_4 + V_1 V_4 and V_8)

Block 2, comprising (say) 6 plots (C_1 to C_4 + V_2 and V_3) and

Block 3, encompassing (say) 7 plots (C_1 to C_4 + V_5, V_6 and V_7)

So as to construct relatively homogenous stratification of area in each block. Thus number of test varieties falling in each block are n = 3, 2 and 3, respectively. Randomize all the entries (check plus varieties) into each block as depicted below:

Field layout : Augment Design II

Block 1 Block 2

C_2	C_4	V_8	C_1	V_1	C_3	V_4	V_3	C_3	C_2
V_2	C_1	C_4	V_7	C_3	C_1	V_5	C_2	V_6	C_4

Block 2 contd... Block 3

c) Statistical analysis: This can be shown with the same data on number of tillers in 8 varieties and 4 checks (replicated 3 times) of linseed as in design 1 but average data arranged block-wise as under Table 6.

Table 6: Number of tillers in 8 varieties and 4 checks of linseed grown in augmented design II.

Checks	Symbolic data					Actual data				
(C_i)	Replication (b_j)			Total	**Mean**	Replication			Total	Mean
	1	2	3	(T_{ci})	(c_i)	1	2	3		
C_1	x_1	x_2	x_3	Tc_1	$\bar{C}_1$	10	8	6	24	8.0
C_2	x_4	x_5	x_6	Tc_2	$\bar{C}_2$	3	5	6	14	4.7
C_3	x_7	x_8	x_9	Tc_3	$\bar{C}_3$	10	9	8	27	9.0
C_4	x_{10}	x_{11}	x_{12}	Tc_4	$\bar{C}_4$	15	14	18	47	15.7
Total (C)	Tcb_1	Tcb_2	Tcb_3	Tc	$\bar{C}$	38	36	38	112	37.3

Test var (Vi) (unadjusted)	Symbolic data				Actual data			
	Replication (b_j)				Replication			
	1	2	3		1	2	3	
	V_8	V_3	V_7		30	4	8	
	V_1	V_2	V_5		6	40	11	
	V_4		V_6		24		18	
Total (V)	T_{vb1}	T_{vb2}	T_{vb3}	T_V	60	44	37	141
Total	T_{b1}	T_{b2}	T_{b3}	GT	98	80	75	253

Table 7: The skeleton of ANOVA for augmented design II

Source of variation	**d.f.**	**SS**	**MSS**	**F**
Blocks	b-1 = 2	bSS	bMS	bMS/EMS
Entries	e-1 = 11	eSS	eMS	eMS/EMS
Checks	c-1 = 3	cSS	cMS	cMS/EMS
Varieties	V-1 = 7	vSS	vMS	vMS/EMS
Checks Vs. varieties	1	cvSS	cvMS	cvMS/EMS
Error	(c-1) (b-1)=6	ESS	EMS	---
Total	N-1 = 19	TSS		

Now different components of ANOVA can be calculated as follows:

However, since only the checks (c) but not the test varieties (v) are replicated in this design, adjustment of means of v must be done before ANOVA with the help of the following effects.

i) Blocks effects (bj) = 1/C (T_{bj} – $\overline{C}$ – T_{vbj}) (j=1 to b)

b_1 = ¼ (98 – 37.3 – 60) = ¼ (0.7) = 0.17

b_2 = ¼ (80 – 37.3 – 44) = ¼ (-1.3) = -0.32

b_3 = ¼ (75 – 37.3 – 37) = ¼ (0.7) = 0.17

Counter check : $\sum_{j=1}^{b} bj = 0$

Ignoring rounding errors. In this case (Σbj = 0.17 - 0.32 + 0.17 = 0)

ii) Mean effect (m) = 1/e (GT – (b-1) $\overline{C}$ - $\sum_{1}^{b} n_j \; b_j$) where, n_j is the number of V occurring in j block.

In the present case

m = 1/12 [(253 – (3-1) 37.3 – (3 x 0.17 + 2 x (-0.32) + 3 x 0.17)]

= 1/12 [253 - 74.6 - 0.38]

= 1/12 x 178.02

= 14.8

If n_j is the same in all blocks, then

m = 1/e (GT – (b – 1) $\overline{C}$)

iii) Check effects ($\overline{C_i}$) = $\overline{C_i}$ – m (i = 1 to C)

C_1 = 8.0 – 14.8 = -6.8; C_2 = 4.7 – 14.8 = -10.1

C_3 = 9.0 – 14.8 = -5.8; C_4 = 15.7 – 14.8 = 0.9

iv) Adjusted mean of test variety (Vi) and genotypic effects (Vi") are now computed as under Table 8.

Table 8 : Adjusted means and genotypic effects.

Test variety	Unadjusted mean (Vi)	Adjusted mean (Vi' = Vi-bj)	Genotypic effect (Vi'' = Vi'-m)
V_1	6	6 – 0.17 = 5.83	5.83 – 14.8 = -8.97
V_2	40	40 + 0.32 = 40.32	40.32 – 14.8 = 25.52
V_3	4	4 + 0.32 = 4.32	4.32 – 14.8 = -10.48
V_4	24	24 – 0.32 = 23.83	23.83 – 14.8 = 9.03
V_5	11	11 – 0.17 = 10.83	10.83 – 14.8 = -3.97
V_6	18	18 – 0.17 = 17.83	17.83 – 14.8 = 3.03
V_7	8	8 – 0.17 = 7.83	7.83 – 14.8 = -6.97
V_8	30	30 – 0.17 = 29.83	29.83 – 14.8 = 15.03

b_j is block effects of block where Vi falls in

ANOVA is now carried out using the data in tables 8 and previous table developed different value as under :

GCF (General correction factor) $GT^2/N = (253)^2/20 = 3200.45$

$$bSS = \sum_{1}^{b} T_{bj}{}^2/(c+n_j) - GCF = \frac{98^2}{7} + \frac{80^2}{6} + \frac{75^2}{7} - 3200.45$$

$$= 1372 + 1066.66 + 803.57 - 3200.45$$

$$= 3242.23 - 3200.45 = 41.78$$

$$bMS = bSS/(b-1) = 41.78/2 = 20.89$$

$$TSS = \sum_{i=1}^{c}\sum_{j=1}^{b} + C_{ij}^2 + \sum_{i}^{v} Vi - GCF$$

$$= x^2{}_1 + x_2{}^2 + \text{---} + x_{12}{}^2 + v_8{}^2 + \text{---} + V_6{}^2 - GCF$$

$$= 10^2+8^2+ 6^2 +\text{---}+ 14^2 + 18^2 + 30^2 + 4^2 +\text{---}24^2+18^2- 3200.45$$

$$= 4897.00 - 3200.45 = 1696.55$$

cCF = (check correction factor) = $Tc^2/bc = 112^2/12 = 1045.3$

$$cSS = \sum_{1}^{c} T_{ci}/b - cCF = \frac{24^2+14^2+27^2+47^2}{3} - 1045.3$$

$$= 1236.6 - 1045.3 = 191.3$$

$$cMS = cSS / (c-1) = 191.3/3 = 63.7$$

$$eSS = (m \times GT) + \left(\sum_{1}^{b} b_j \; T_{bj}\right) + \left(\sum_{1}^{c} T_{Ci} \; Ci\right) + \left(\sum_{1}^{v} V_i \; V''\right) - \left(\sum_{1}^{b} T_{bj}^2/(c+n_j)\right)$$

$$= (14.8 \times 253) + (0.17 \times 98) + (-0.32 \times 80) + (0.17 \times 75) +$$
$$[24 \times (-6.81)] + [14 \times (-10.1)] + (27 \times 5.8) + (4.7 \times 9) +$$
$$[6 \times (-8.97)] + (40 \times 25.52) + [4 \times (-10.48)] + [24 \times 9.03)] +$$
$$[11 \times (-3.97)] + (18 \times 3.03) + [8 \times (-6.97)] + [30 \times 15.03)]$$
$$-\left(\frac{98^2}{4+3} + \frac{80^2}{4+2} + \frac{75^2}{4+3}\right) = 1634.8$$

$eMS = eSS / (e-1) = 1634.1 / 11 = 148.6$

vCF (variety correction factor) $= Tv^2 / V = 141^2 / 8 = 2485.1$

$$vSS = \sum_{1}^{v} Vi^2 \quad vCF = 6^2 + 40^2 + ... + 8^2 + 30^2 - 2485.1$$

$= 3637.0 - 3485.1 = 1151.9$

$VMS = VSS / (v-1) = 1151.9/7 = 164.5$

$cvSS = eSS - cSS - vSS = 1634.8 - 191.3 - 1151.9 = 291.6$

$ESS = TSS - eSS - bSS = 1696.55 - 1634.80 - 41.78 = 19.97$

$EMS = ESS / (c-1)(b-1) = 19.97/6 = 3.32$

Presentation of results: The ANOVA table is presented as under:

Table 9 : ANOVA

Source of variation	d.f.	SS	MSS	F
Blocks (b)	2	41.78	20.89*	6.29
Entries (e)	11	1634.80	148.60**	44.75
Checks (c)	3	191.30	63.70**	19.18
Varieties (v)	7	1151.90	164.50**	49.54
Checks Vs. varieties (cvs.v)	1	291.60	291.60**	87.88
Error	6	19.97	3.32	
Total		1696.55		

*, ** -$p<0.05$, $p<0.01$ respectively

Mean performance is same as in Table 5

Inferences: Same as in augmented design I, except that here variance due to block effect is significant ($P<0.05$). Now, following four standard errors are helpful in drawing conclusions:

SEd_1 (between any two check means) $= \sqrt{\frac{2EMS}{b}} = \sqrt{\frac{2\times 3.32}{3}}$

$= \sqrt{2.213} = 1.48$

SEd_2 (between any two means of test varieties) in the same block $= \sqrt{2EMS} = \sqrt{2\times 3.32} = 2.57$

SEd_3 (between any two test varieties not in the same block) =

$$\sqrt{2EMS\left(1+\frac{1}{c}\right)} = \sqrt{2\times 3.32\times\left(1+\frac{1}{4}\right)}$$

$$= \sqrt{6.64\times 1.25} = \sqrt{8.3} = 2.88$$

SEd_4 (between mean of a check and a test variety)

$$= \sqrt{EMS\left(1+\frac{1}{b}+\frac{1}{c}+\frac{1}{bc}\right)} = \sqrt{3.32\times 1.66} = 2.34$$

CD is than calculated using these SEd x t at 5% or 1% levels (of d.f. = 6) as the case may be :

Table 10: ANOVA of augmented design II for large collections

(N = 1300, b=10, C=10, V=1200, e=1210)

Source of variation	**d.f.**	**MSS**
Block (b)	b –1 = 9	
Entries (e)	e – 1 = 1209	
Checks (c)	c – 1 = 9	
Test varieties (v)	v – 1 = 1199	
c Vs. v	1	
Error	(c –1) (b –1) = 81	
Total	(N-1) = 1299	

Extension of augmented design II for large collections

Field designs layout and ANOVA etc. all will follow the same pattern as elaborated above. For a testing of large collections (e.g. 1200 genetic stocks + 10 checks replicated 10 times) the following ANOVA is exemplified (Table 10).

General consideration: Some general points in connection with greater utility of augmented design II are:

1. The number of test genotypes may be kept the same in each block. This will greatly facilitate statistical analysis.
2. A block may continue in next (contiguous) strip also. However, as far as possible, plots in each block may be kept homogeneous for soil factors. Homogeneity of soil in a block can thus be maintained by increasing the number of blocks (adequate stratification of experimental area) and by reducing the plot size.
3. Augmented design II serves as the best design for initial screening of large collection generally unmanageable by any other design. The choicest genotypes are reevaluated in RBD for intensive investigation.
4. While determining genetic divergence among extensively large number of genotypes through multivariate analysis, this design can help both evolutionary geneticists and plant breeders alike.

Thus, augmented designs are unique in that they are employed in situation where no other design can logically be adopted.

Augmented BIB Design

Consider any BIB design with parameters v, b, r, k and λ and augment every block by adding s-th (s=1, 2.....,t) supplementary treatment to β_s plots and call the resultant design as D*. It will have (v+t) treatments and every block will have $k^* = k + \sum_s \beta_s = k + k_0$ plot.

Example: The layout plan of a BIB augmented design is given along with the yield per plot. Analyse data.

Yield in kg/plot

(3) 25.3	(6) 19.9	(9) 29.0	(11) 24.6	(14) 20.5
(3) 23.0	(4) 19.8	(8) 33.3	(12) 22.7	(14) 18.6
(10) 16.2	(11) 19.3	(12) 31.7	(13) 26.6	(14) 20.7
(2) 27.3	(5) 27.0	(8) 35.6	(11) 17.4	(14) 18.7
(7) 23.4	(8) 30.5	(9) 30.8	(10) 32.4	(14) 22.0
(4) 30.6	(5) 32.4	(6) 27.2	(10) 32.8	(14) 19.3
(1) 34.7	(5) 31.1	(9) 25.7	(12) 30.5	(14) 24.1
(3) 34.4	(5) 32.4	(7) 33.3	(13) 36.9	(14) 17.5
(1) 38.2	(2) 32.9	(3) 37.3	(10) 31.3	(14) 21.2
(2) 28.7	(4) 30.7	(9) 26.9	(13) 35.3	(14) 23.5
(1) 36.6	(4) 31.1	(7) 31.1	(11) 28.4	(14) 21.3
(1) 31.8	(6) 33.7	(8) 27.8	(13) 41.1	(14) 20.7
(2) 30.3	(6) 31.5	(7) 39.3	(12) 26.7	(14) 19.8

The figures in bracket indicate treatment numbers.

Calculations

i) Here $v^* = v + t = 14$, $b = 13$, $r = 4$ for basic treatments and 13 for supplementary treatment and $k^* = k + 1 = 5$.

ii) Computation of Q_i's

Table

Treatment number	**Treatment total (T_i)**	**Block numbers in which i-th treatment occurs**	**Sum of all the blocks containing i-th treatment ($\sum_{(i)} B_j$)**	$Q_i = T_i - \sum_{(i)} B_j / k^*$
1	141.3	7, 9, 11, 12	610.5	19.20
2	119.2	4, 9, 10, 13	579.6	3.28
3	120.2	1, 2, 8, 9	552.1	9.58
4	112.2	2, 6, 10, 11	553.2	1.56
5	122.9	4, 6, 7, 8	568.9	9.12
6	112.3	1, 6, 12, 13	564.3	-0.56
7	127.1	5, 8, 11, 13	589.6	9.18
8	127.2	2, 4, 5, 12	537.6	19.68
9	112.4	1, 5, 7, 10	549.5	2.50

Treatment number	Treatment total (T_i)	Block numbers in which i-th treatment occurs	Sum of all the blocks containing i-th treatment ($\sum_{(i)} B_j$)	$Q_i = T_i - \sum_{(i)} B_j / k^*$
10	112.7	3, 5, 6, 9	556.8	1.34
11	89.7	1, 3, 4, 11	508.2	-11.94
12	111.6	2, 3, 7, 13	525.2	6.48
13	139.9	3, 8, 10, 12	569.2	26.06
14	267.8	1 to 13	1816.3	-95.46

$$\sum_i Q_i^2 = 190045, \qquad \mu^* = \frac{r-\lambda}{rk^*} = \frac{3}{20}$$

Treatment S.S. (adjusted)

$$= \frac{1}{r(1-\mu^*)} \sum_i Q_i^2 - \frac{\mu^*}{(1-\mu^*)rv} Q^2 + \sum_s \frac{Q_{ts}^2}{b\beta_s} = 1229.00,$$

Here $Q_. = \sum_{i=1}^{v} Q_i$

Total S.S. = 2475.28

Block S.S. (unadjusted) = 571.43

Q_{ts} denote the adjusted total of s-th (s=1, 2..... t) supplementary treatment.

iii) Analysis of variance

ANOVA Table

Source	d.f.	S.S.	M.S.	F
Blocks (unadjusted)	12	571.43	47.62	
Treatments (adjusted)	13	1229.00	94.54	5.46**
Error	39	674.85	17.30(s_e^2)	
Total	64	2475.28		

** indicates significant at 1 per cent level

It can be inferred that the treatment effects significantly

iv) Variances of elementary contrasts

Variance of elementary contrasts are

$$V\left(\hat{t}_i - \hat{t}_j\right) = \frac{2s_e^2}{r(1-\mu^*)} = 10.18 \text{ kg}^2$$ when i and j are treatments of basic set

$$= \left[\frac{1}{r(1-\mu^*)}\left(1-\frac{\mu^*}{r(v)}\right)+\frac{1}{b\beta_i}\right]s_e^2$$

= 6.36 kg^2 for comparison between basic and supplementary treatments.

❐ ❐ ❐

Chapter – 16

Block Designs

Block designs are widely used in many fields of research. Their most common type is the randomized block design. It has many advantages over other designs, but it is not applicable in all circumstances. If the number of treatments is too large to preserve homogeneous conditions within complete block or if the size of block is determined by the nature of experimental material or technique, we have to resort to incomplete block design where every treatment is not included in a given block.

Balanced Incomplete Block Designs (BIBD)

The most important class of incomplete block designs is the balanced incomplete block (BIB) design introduced by Yates (1936). A BIB design is an arrangement of v treatments into b blocks each of size k(<v) treatments, satisfying the followings conditions.

(i) Every treatment occurs at most once in each block.

(ii) Every treatment occurs in exactly r blocks.

(iii) Every pair of treatments occurs together in exactly λ blocks.

These designs have the characteristic that all the contrasts are confounded with the blocks to the same extent and every elementary contrast is estimated with same variance.

Partially Balanced Incomplete Block Design (PBIBD)

The restriction that each pair of treatments should occur exactly λ times in the design is demanding and unfortunately, the BIB designs do not exist for all meaningful parametric combinations. Moreover, these designs require a large number of replicates for given v and k. This led to the investigation of other types of incomplete block designs in which the restriction of every contrast being confounded with blocks to same extent is relaxed. The design where some of the treatment contrasts are more confounded than others are partially balanced incomplete block (PBIB) designs introduced by Bose and Nair (1939). For PBIB designs with m-associate classes, Var (a_i-a_j) may take m different values depending upon whether i^{th} and j^{th} treatments are first, second or m^{th} associates. Bose and Shimamoto (1952) introduced the concept of partially balanced association scheme and the definition of PBIB designs was rephrased in this context. The definition of m class-association scheme and the PBIB design is given below:

Given v treatments, 1, 2, ..., v, a relation satisfying the following conditions is said to be an association scheme with m classes:

(i) Any two treatments are either 1^{st}, 2^{nd}, ..., or m-th associates, the relation of association being symmetrical; that is, if the treatment θ is i-th associate of ϕ, then ϕ is also i-th associate of θ.

(ii) Each treatment θ has n_i i-th associates, the number n_i being independent of the particular treatment θ chosen.

(iii) If any two treatments θ and ϕ are i-th associates, then the number of treatments that are j-th associates of θ, and k-th associates of ϕ is p^i_{jk} and is independent of the pair of i-th associate θ and ϕ.

The numbers v, n_i (i=1, 2, ..., m) and p^i_{jk} (i, j, k=1, 2,.....,m) are called the parameters of association scheme.

Given an m-class association scheme, we have a PBIB design with m classes, based on the scheme; if it is possible to arrange the v

treatments into b blocks such that.

(i) Each block contains k distinct treatments.

(ii) Each treatment occurs in r blocks.

(iii) Any two treatments which are i-th associates occur together in λ_i blocks, (i=1, 2,...., m), where λ_i is constant independent of the pair of the i-th associates chosen.

The integers v, b, r, k, λ_i (i= 1, 2, ..., m) are called the parameters of the PBIB design.

From practical view point, the two associates class PBIB designs are important and an exhaustive list of these design was prepared by Bose, Clatworthy and Shrikhande (1954). The list was subsequently revised by Clatworthy (1956, 1973). Bose and Shimamoto (1952) have classified the two associates class PBIB designs in to five following categories depending upon the association scheme on which they are based.

i) Group divisible (GD)

ii) Simple (SL)

iii) Triangular (T_2)

iv) Latin square type (L_i)

(v) Cyclic (C)

Partially Efficiency Balanced (PEB) Design

In BIB and PBIB designs, the restriction of equal block size and equal number of replications for each treatment may be a serious practical obstacle in many possible experimental circumstances in biological research. With biological material there are natural units that can be used as block and these blocks contain varying number of units which are not under the control of experimenter. This may be clear from the following example.

(i) In a nutritional experiment with pigs, the number of young ones in the available liters may be unequal, being in some cases equal to the number of treatments but in others smaller or larger.

(ii) A psychologist have to use families as blocks, but, the numbers of family members may very from family to family.

(iii) The insecticidal sprays, each of two concentrations, are to be compared on different batches of insects. The conditions allow only 4 batches to be sprayed on one day, and the experimenter wants to have 2 replications for each of the 6 treatments and as many as 5 for the control.

(iv) Many new varieties or strains of a crop together with one or two standard varieties are to be compared in one field experiment. The number of plots for the new varieties is usually severely limited by available amount of seed, but no such limitation is there for standard varieties. One could replicate standard varieties more than newly evolved varieties.

A further inconvenience in the use of BIB and PBIBD is that these are not available for every, meaningful parametric combinations. For example, no two associate PBIB design is available in 2 or 3 replications for 18 varieties. Moreover, the designs may not be available for the number of treatments in which the experimenter is actually interested or the available designs may use a large number of replication which the experimenter may not be able to afford.

A partial answer to this problem is provided by the work of Puri and Nigam (1977). They introduced Partially Efficiency Balanced Designs which are available in unequal number of replications and block sizes.

A block Design D (v, b, $\underline{r}$, $\underline{k}$) with v treatments arranged in b blocks of sizes k_1, k_2.... k_b such that i-th treatment be replicated r_i times (i=1, 2...v) is said to be a partially efficiency balanced (PEB) design with m-efficiency classes if a complete set of (v-1) basic contrasts can be partitioned in to m-disjoint classes such that efficiency factor associated with every contrast of i-th class is $E_i = (1-\mu_i)$, where μ_i (i=1, 2...m) is the eigen value of $\underline{M_o}$ matrix with multiplicity ρ_i such that $\sum_{i=1}^{m} \rho_i = v-1$. For PEB design $\underline{M_0}$ will have

spectral decomposition.

$M_0 = \sum_{i=1}^{m} \mu_i \underline{L}_i$ where $\underline{L}_i$ are mutually orthogonal idempotent matrices satisfying $\sum_{i=1}^{m} \underline{L}_i = \underline{I}_v$

The parameters of PEB design with m efficiency classes can be written as v, b, $\underline{r}$, $\underline{k}$, μ_i, ρ_i, $\underline{L}_i$, i =1, 2,...m.

Here

N : (n_{ij}), a v x b incidence matrix of treatment in to blocks associated with every block design D, n_{ij} being the number of times the i-th treatment occurs in the j-th block.

Let $\underline{r} = [r_1, r_2, \ldots\ldots r_v]^T$ is a v x 1 vector of replication with i-th element as r_i, the replication of i-th treatment , i=1, 2, ..., v.

$\underline{k} = [k_1, k_2, \ldots\ldots, k_b]^T$ is a b x 1 vector of block sizes with j-th element as k_j, size of j-th block, j = 1, 2....., b

$\underline{R}$: v x v diagonal matrix with diagonal elements as $\underline{r}$

$\underline{K}$: b x b diagonal matrix with diagonal elements as $\underline{k}$

$\underline{R}^{-1}$:Inverse $\underline{R}$

$\underline{K}^{-1}$: Inverse of $\underline{K}$

n : Total numbers of observation in the design.

T: Stands for Transpose sign.

$\underline{C}$: $= \underline{R} - \underline{N}\,\underline{K}^{-1}\,\underline{N}^T$

$\underline{M}_0 = \underline{R}^{-1}\,\underline{N}\,\underline{K}^{-1}\,N^T - \underline{1}\,\underline{r}^T/n$

A particular class of PEB designs where μ_i takes only two values μ and 0 with multiplicities ρ and (v-ρ-1) is termed as simple PEB design. The parameters of simple PEB designs can be written as.

v, b, $\underline{r}$, $\underline{k}$, μ, ρ, $\underline{L}$

If $\mu_i = \mu$ for all i, the design is called efficiency balanced (EB) [Puri and Nigam (1975) & William (1975)].

Optimality of Designs

In the early stags of agricultural experimentation, only a few experimental designs were at one's disposal. But now-a-days an experimenter can use a variety of experimental designs. Hence whenever the conditions of an experiment allow the possibility of simultaneous existence of a number of designs, the questions of selection of an appropriate design, a design which is easy to analyse and which has some optimum properties, naturally arises. A systematic study of the specification of optimum experimental designs was under taken by Kiefer (1958, 1959) where he introduced the various optimality criteria (A, D, E, L, M) and discussed interrelations amongst these and established the optimality property of some well known designs.

A-optimum Designs

One of the objectives of designing experiments is to ensure enhanced accuracy and reliability of conclusions and inferences. Such accuracy depends on variances of estimates of treatment contrasts or comparisons, the less such variance the more is the accuracy. Designs can be compared by using suitable functions of variances of estimates of all normalized independent (orthogonal) contrasts of treatment effects. One such function is the mean of reciprocals of variances of estimates of each of the normalized contrasts of treatments when the contrasts are mutually orthogonal. A design where such a mean is maximum among all the competing designs is called an A-optimum design. Mathematically mean of the reciprocals of the Eigen value of C matrix of a design as a criterion for judging a design as A-optimal.

D-optimum Designs

In A-optimum designs discussed above, we have used the mean of the reciprocals of the Eigen value of C Matrix of a design as a criterion for judging a design. In D-optimum designs, the criterion used is the product of the Eigen values and this should be minimum

for a D-optimum design over similar competing designs. It is not possible to indicate what criterion should be used i.e. whether D-optimum design or A-optimum design for discriminating designs. For any pronouncement in this regard, the situation and objectives of the experiments have to be taken into account.

E-optimum design

Again, there is one more of type of optimum design known as E-optimum designs. A design is said to be E-optimum if the maximum Eigen value of C Matrix is less than the maximum Eigen value of C Matrix of any other design.

This way, there are several other criterion for optimization and considerable research activities are available in literature in regard to such optimum designs. However, no systematic tabulation of such designs exists in literature. For the definition of L and M optimality criteria refer to Kiefer (1958).

There can be another type of optimum designs. The designs discussed above are assessed by taking into account, total information that a design can provide (emanating from all orthogonal contrasts). But there are situations where experimenters are interested mainly in certain specific contrasts. If there can be designs which allow estimation of such contrasts with full precision such designs are also optimum though in a limited but more meaningful manner.

Analysis of Two Associates Partially Balance Incomplete Block Design Analysis of some important types of two associate PBIB designs are given. The procedure of analysis is very much simple and straight forward.

Group divisible (GD) design

A PBIB with two associate class is said to be GD it there are v=mn treatments and the treatments can be arranged in m groups of n treatment each, such that any two treatment of the same group are first associate and any two treatments from different groups are

second associates. The parameters of GD design are v = mn, b, r, k, λ_1 and λ_2, $n_1 = (n-1)$, $n_2 = n(m-1)$

Example: The following gives the plan and yield in kg/plot of 15 varieties of wheat in yield trial. The figures in bracket indicate the variety number.

Yield in kg/plot

(15) 2.4	(9) 2.5	(1) 2.6	(13) 2.0
(5) 2.7	(7) 2.8	(8) 2.4	(1) 2.7
(10) 2.6	(1) 2.8	(14) 2.4	(2) 2.2
(15) 3.4	(11) 3.1	(2) 2.1	(3) 2.3
(6) 4.1	(15) 3.3	(4) 3.3	(7) 2.9
(12) 3.4	(4) 3.2	(3) 2.8	(1) 3.0
(12) 3.2	(14) 2.5	(15) 2.4	(8) 2.6
(6) 2.3	(3) 2.3	(14) 2.4	(5) 2.7
(5) 2.8	(4) 2.8	(2) 2.6	(13) 2.5
(10) 2.5	(12) 2.7	(13) 2.8	(6) 2.6
(9) 2.6	(7) 2.6	(10) 2.3	(3) 2.4
(8) 2.7	(6) 2.7	(2) 2.5	(9) 2.6
(5) 3.0	(9) 3.6	(11) 3.2	(12) 3.2
(7) 3.0	(13) 2.8	(14) 2.4	(11) 2.5
(10) 2.4	(4) 2.5	(8) 3.2	(11) 3.1

Analyse the data and draw conclusions.

Calculations

(i) Here v = 5x3, b= 15, r = k = 4 $\lambda_1 = 0$, $\lambda_2 = 1$ and also m = 5, n = 3, $n_1 = 2$, $n_2 = 12$.

The key to the design is the following association scheme among the treatment number.

	1	6	11
Groups	2	7	12
	3	8	13

4	9	14
5	10	15

(ii) Calculation of Q_{ij}'s

Computation of Q_{ij}'s

Treatment number (s,j)	Treatment total T_{sj}	Blocks in which (s,j) th treatment occurs	S_{sj}	$Q_{sj}=T_{sj}-S_{sj}/k$
(1,1)	11.1	1, 2, 3, 6	42.5	0.475
(2,1)	9.4	3, 4, 9, 12	42.1	-1.125
(3,1)	9.8	4, 6, 8, 11	42.9	-0.925
(4,1)	11.8	5, 6, 9, 15	47.9	-0.175
(5,1)	11.2	2, 8, 9, 13	44.0	0.200
(1,2)	11.7	5, 8, 10, 12	44.4	0.600
(2,2)	11.3	2, 5, 11, 14	44.8	0.100
(3,2)	10.9	2, 7, 12, 15	43.0	0.150
(4,2)	11.3	1, 11, 12, 13	42.9	0.575
(5,2)	9.8	3, 10, 11, 15	41.7	-0.625
(1,3)	11.9	4, 13, 14, 15	45.8	0.450
(2,3)	12.5	6, 7, 10, 13	46.7	0.825
(3,3)	10.1	1, 9, 10, 14	41.5	-0.275
(4,3)	9.7	3, 7, 8, 14	41.1	-0.575
(5,3)	11.5	1, 4, 5, 7	44.7	0.325

Here T_{sj} denote the total of all the plots receiving (s,j) treatment. S_{sj} denote that total value of all those blocks which receive the (s,j) treatment.

We now arrange the treatments group wise and propose the following two way Table.

Group	1	2	3	θs.
1	0.475	0.600	0.450	1.252
2	-1.125	0.100	0.825	-0.200
3	-0.925	0.150	-0.275	-1.050
4	-0.175	0.575	-0.575	-0.175
5	0.200	-0.625	0.325	-0.100

$$\sum_s \sum_j Q^2_{sj} = 4.5262 \qquad \sum_s Q^2 s = 3.5087$$

We compute the following

$$\mu_1 = \frac{r-\lambda_1}{k} = \frac{1}{4} \qquad \mu_2 = \frac{rk - v\lambda_2}{rk} = \frac{1}{16}$$

Treatment SS (adjusted) =

$$\frac{1}{r(1-\mu_1)}\left(\sum_s \sum_j Q^2_{sj} - \sum_s \frac{Q^2 s}{n}\right) + \frac{1}{r(1-\mu_2)} \sum_s \frac{Q^2 s}{n} = 1.5640$$

Block S.S. unadjusted = 4.9233 Total S.S. = 9.1733

ANOVA

Source	D.F.	S.S.	M.S.	F
Block (unadjusted)	14	4.9233		
Treatment (adjusted)	14	1.5640	0.1117	1.289
Error	31	2.6860	0.0866	
Total	59	9.1733		

From the above table it is inferred that the effects of varieties do not differ significantly. Variance of elementary contrasts can be worked out as:

$$V(t_s - t_j) = \frac{2s_e^2}{r(1-\mu_1)} = 0.0577 \text{ when s and j are first associate}$$

$$= \frac{2s_e^2}{r(1-\mu_1)}\left(1 + \frac{\mu_2 - \mu_1}{n(1-\mu_2)}\right) = 0.0539 \text{ otherwise}$$

L_2 PBIB Design

In L_2 type PBIB designs, there are v^2 treatment arranged in a v x v square so that any two treatments are first associate if they occur in same row or same column and are second associate otherwise. The parameters of L_2 PBIB design are v^2, b, r, k, λ_1, λ_2, $n_1 = 2(v-1)$ and $n_2 = (v-1)^2$.

Example

The layout plan of PBIB design 6 blocks of size 3 is given along with the yield/plot. Analyse the data and draw conclusions.

Yield in kg/plot		
(1) 31	(2) 28	(3) 25
(4) 61	(5) 61	(6) 54
(7) 46	(8) 56	(9) 59
(1) 48	(4) 61	(7) 42
(2) 35	(5) 61	(8) 59
(3) 53	(6) 62	(9) 46

Calculation

Here $v^2 = 9$, b=6, r = 2, k = 3, $\lambda_1 = 1$, $\lambda_2 = 0$, $n_1 = 4$, $n_2 = 4$

The key to the design is the following association scheme among the treatment number.

1	2	3
4	5	6
7	8	9

Computation of Q_{ij}'s

Treatment number (s,j)	Treatment total T_{sj}	Blocks in which (s,j) th treatment occurs	S_{sj}	$Q_{sj}=T_{sj}-S_{sj}/k$
(1,1)	79	1, 4	235	0.67
(1,2)	63	1, 5	239	-16.67
(1,3)	84	1, 6	251	0.33
(2,1)	122	2, 4	327	13.00
(2,2)	122	2, 5	331	11.67
(2,3)	116	2, 6	343	1.67
(3,1)	88	3, 4	312	-16.00
(3,2)	115	3, 5	316	9.67
(3,3)	105	3, 6	328	-4.33

We now arrange the treatments in v x v square and prepare the following two-way Table.

Group	1	2	3	$Q_{i.}$
1	0.67	-16.67	0.33	-15.67
2	13.00	11.67	1.67	26.34
3	-16.00	9.67	-4.33	-10.66
Q_j	-2.33	4.67	-2.33	

$$\sum_i \sum_j Q^2_{ij} = 954.6823$$

$$\sum_i Q^2 i. = 1052.9801$$

$$\sum_j Q^2 .j = 32.6667$$

The following can be easily computed

$$\mu_1 = \frac{(r-\lambda_1)+(v-1)\ (\lambda_1-\lambda_2)}{rk} = \frac{1}{2}$$

$$\mu_2 = \frac{(r-2\lambda_1+\lambda_2)}{rk} = 0$$

Treatment SS (adjusted)

$$\frac{1}{rv\,(1-\mu_1)}\left(\sum_i Q^2_{i.} - \sum_j Q^2_{.j}\right) + \frac{1}{r\,(1-\mu_2)}$$

$$\left(\sum_i \sum_j Q^2_{ij} - \frac{1}{v}\sum_j Q^2_{i.} - \frac{1}{v}\sum_j Q^2_{.j}\right) = 658.2823$$

Block S.S. (unadjusted) = 1820.67

Total S.S. = 2716.00

ANOVA

Source	D.F.	S.S.	M.S.	F
Block (unadjusted)	5	1820.67	364.13	
Treatment (adjusted)	8	658.28	82.29	1.39
Error	4	237.05	29.26	
Total	17	2716.00		

It is concluded that the treatment effect do not differ significantly.

$$v(\hat{t}_i - \hat{t}_j) = \frac{2}{rv}\left(\frac{1}{(1-\mu_1)} + \frac{(v-1)}{(1-\mu_2)}\right)s_e^2 \text{ if i and j are first}$$

associate = 79.01

$$= \frac{2}{rv}\left(\frac{2}{(1-\mu_1)} + \frac{(v-2)}{(1-\mu_2)}\right)s_e^2 \text{ otherwise} = 98.77$$

Augmented PBIB Designs

Consider any m-associate PBIB design, D, with parameters v, b, r, k and λ_i (i=1, 2, ……; m). Suppose we add t supplementary treatments by adding s-th (s=1, 2, …. , t) supplementary treatment to β_s plots $\left(\sum_s \beta_s = k_0\right)$ of every block and call the resultant design as D*. It is clear the k*, the block size of the new design, will be $k+k_0$ and v*, the number of treatments in the new design, will be v+t.

Augmented GD Design:

Consider a GD design with parameters v=mn, b, r, k, λ_1 and λ_2 and obtain D* by augmenting t new treatments as explained above. The number of treatments and the block size of D* will be mn+t and $k+k_0$ respectively.

Example: Following table gives the plan and yield in kg. per plot of 16 varieties of Wheat in a particular yield trial. The figures in bracket indicate variety number.

Yield in kg./plot

(15) 2.4	(9) 2.5	(1) 2.6	(13) 2.0	(16) 2.3
(5) 2.7	(7) 2.8	(8) 2.4	(1) 2.7	(16) 2.4
(10) 2.6	(1) 2.8	(14) 2.4	(2) 2.2	(16) 2.7
(15) 3.4	(11) 3.1	(2) 2.1	(3) 2.3	(16) 2.9
(6) 4.1	(15) 3.3	(4) 3.3	(7) 2.9	(16) 3.1
(12) 3.4	(4) 3.2	(3) 2.8	(1) 3.0	(16) 2.7
(12) 3.2	(14) 2.5	(15) 2.4	(8) 2.6	(16) 2.4
(6) 2.3	(3) 2.3	(14) 2.4	(5) 2.7	(16) 2.5

(5) 2.8	(4) 2.8	(2) 2.6	(13) 2.5	(16) 2.8
(10) 2.5	(12) 2.7	(13) 2.8	(6) 2.6	(16) 2.9
(9) 2.6	(7) 2.6	(10) 2.3	(3) 2.4	(16) 3.0
(8) 2.7	(6) 2.7	(2) 2.5	(9) 2.6	(16) 2.1
(5) 2.7	(9) 3.6	(11) 3.2	(12) 3.2	(16) 2.6
(7) 3.0	(13) 2.8	(14) 2.4	(11) 2.5	(16) 2.3
(10) 2.4	(4) 2.5	(8) 3.2	(11) 3.1	(16)

Analyse the data and draw conclusion

Calculations

(i) Here $v^* = v + t = 16$, $b = 15$, $r = 4$ $k^* = k+k_0 = 5$

If Q_{ij} are the adjusted totals of (i, j) treatment of a GD association schemes, then we have

(ii) Calculation of Q_{ij}'s

Computation of Q_{ij}

Treatment number (i, j)	Treatment tptal (T_{ij})	Block numbers in which (i, j) treatments occur	Sum of blocks where (i, j) treatments occur (S_{ij})	$Q_{ij} = T_{ij} - S_{ij}/k^*$
(1, 1)	11.1	1, 2, 3, 6	52.6	0.58
(2, 1)	9.4	3, 4, 9, 12	52.6	-1.12
(3, 1)	9.8	4, 6, 8, 11	54.0	-1.00
(4, 1)	11.8	5, 6, 9, 15	58.8	0.04
(5, 1)	11.2	2, 8, 9, 13	53.7	0.46
(1, 2)	11.7	5, 8, 10, 12	55.0	0.70
(2, 2)	11.3	2, 5, 11, 14	55.9	0.12
(3, 2)	10.9	2, 7, 12, 15	52.2	0.46
(4, 2)	11.3	1, 11, 12, 13	52.3	0.84
(5, 2)	9.8	3, 10, 11, 15	526	-0.72
(1, 3)	11.9	4, 13, 14, 15	55.6	0.78
(2, 3)	12.5	6, 7, 10, 13	56.7	1.16
(3, 3)	10.1	1, 9, 10, 14	52.1	-0.32
(4, 3)	9.7	3, 7, 8, 14	51.3	-0.56
(5, 3)	11.5	1, 4, 5, 7	55.4	0.42
16*	38.7	1 to 15	202.7	-1.84

* Indicates supplementary treatment

We arrange the treatments groupwise and prepare the following two-way table:

Values of Q_{ij}

Group	1	2	3	$Q_{i.}$
1	0.58	0.70	0.78	2.06
2	-1.12	0.12	1.16	0.16
3	-1.00	0.46	-0.32	-0.86
4	0.04	0.84	-0.56	0.32
5	0.46	-0.72	0.42	0.16

$Q_{..} = 1.84,$

$\sum_i \sum_j Q_{ij}^2 = 7.2904,$

$\sum_i Q_{i.}^2 = 5.1368$

The following computations can now easily be performed

$$A_1^* = \frac{1}{r}\left[\sum_i \sum_j Q_{ij}^2 - \frac{1}{n}\sum_i Q_{i.}^2\right] = 1.3945$$

$$A_2^* = \frac{1}{rn}\left[\sum_i Q_{i.}^2 - \frac{Q_{..}^2}{m}\right] = 0.6771$$

$$\mu_1^* = \frac{vr}{bk^*}\mu_1 = \frac{1}{20}, \text{ where } \mu_1 = \frac{(r-\lambda_1)}{rk} = \frac{1}{16}$$

$$\mu_2^* = \frac{vr}{bk^*}\mu_2 = \frac{1}{5}, \text{ where } \mu_2 = \frac{(rk - v\lambda_2)}{rk} = \frac{1}{4}$$

$$\text{Treatment S.S. (adjusted)} = \frac{Q_{..}^2}{rv} + \sum_{i=1}^{2} \frac{A_i^*}{1-\mu_1^*} + \sum_s \frac{Q_{ts}^2}{b\beta_s}$$

$$= 2.5964$$

Total S.S. = 10.97,

Block S.S. (unadjusted) = 4.39

iii) Analysis of variance

ANOVA Table

Source	d.f.	S.S.	M.S.	F
Blocks (unadjusted)	14	4.39	0.3135	
Treatments (adjusted)	15	2.60	0.1733	2.053**
Error	45	3.98	0.0844	
Total	74	10.97		

** indicates significance at 5 per cent level

iv) Variance of elementary contrasts

Variance of elementary contrasts are

$$V\left(\hat{t}_i - \hat{t}_j\right) = \frac{2}{r(1-\mu^*)} s_e^2 = 0.0444 \text{ kg}^2$$ when i and j are first associates

$$= \frac{1}{r(1-\mu_1^*)}\left[1+\frac{\mu_2^*-\mu_1^*}{n(1-\mu_2^*)}\right] s_e^2 = 0.0472 \text{ kg}^2$$ when i and j are second associates

$$= \left[\frac{1}{vr}+\frac{n-1}{m(1-\mu_1^*)}+\frac{m-1}{vr(1-\mu_2^*)}+\frac{1}{b\beta_j}\right] s_e^2 = 0.0331 \text{ kg}^2.$$

Where one is supplementary and the other is basic treatment

Augmented L_2 PBIB design:

Consider an L_2 PBIB design with parameters v^2, b, r, k λ_i and augment every block by adding s-th (s=1, 2, ..., t) supplementary treatment to β_s plots, and call the resultant design ad D*. It is clear that k*, the block size of the new design, will be $k+k_0$ and v*, the number of treatments in the new design, will be v^2+t.

Example: The layout of PBIB (augmented) design is given alongwith the yield in kg. per plot. The figures in bracket indicate treatment numbers.

Yield in kg./plot

(1) 31	(3) 28	(3) 25	(10) 24
(4) 61	(5) 61	(6) 54	(10) 30
(7) 46	(8) 56	(9) 59	(10) 28
(1) 48	(4) 61	(7) 42	(10) 29
(2) 35	(5) 61	(8) 59	(10) 32
(3) 59	(6) 62	(9) 46	(10) 39

Analyse the data and draw conclusions.

Calculations:

i) Here $v^* + t = 10$, $b = 6$, $r = 2$, $k^* = k + k_0 = 4$

$\lambda_1 = 1$, $\lambda_2 = 0$

ii) If Q_{ij} is the adjusted total of (i, j) treatment of an L_2 association scheme, then we have the following computations:

Table: Computation of Q_{ij}

Treatment number (i, j)	Treatment total (T_{ij})	Block numbers in which (i, j) treatments occur	Sum of blocks where (i, j) treatments occur (S_{ij})	$Q_{ij} = T_{ij} - S_{ij}/k^*$
(1, 1)	79	1, 4	288	7.00
(1, 2)	63	1, 5	295	-10.75
(1, 3)	84	1, 6	314	5.50
(2, 1)	122	2, 4	386	25.50
(2, 2)	122	2, 5	396	23.75
(2, 3)	116	2, 6	412	13.00
(3, 1)	88	3, 4	369	-4.25
(3, 2)	115	3, 5	376	21.00
(3, 3)	105	3, 6	395	6.25
10*	182	1, 2, 3, 4, 5, 6	1076	-87.00

Indicates supplementary treatment

iii) Arrange the 9 treatments in the form of an L_2 association scheme and prepare the following table:

i	j			
	1	2	3	$Q_{i.}$
1	7.00	-10.75	5.50	1.75
2	25.50	23.75	13.00	62.25
3	-4.25	21.00	6.25	23.00
$Q_{.j}$	28.25	34.00	24.75	87.00

$$\sum_i \sum_j Q_{ij}^2 = 2076.25, \quad \sum_i Q_{i.}^2 = 4407.12,$$

$$\sum_j Q_{.j}^2 = 2566.62, \quad Q_{..} = 87.00$$

We then compute the following

$$\mu_1^* = \frac{rv}{bk^*}.\mu_1 = \frac{3}{8} \text{ where } \mu_1 = [(r - \lambda_1) + (v - 1)(\lambda_1 - \lambda_2)]/rk = \frac{1}{2}$$

$$\mu_2^* = \frac{rv}{vk^*}.\mu_2 = 0$$

where $\mu_2 = [(r - 2\lambda_1 + \lambda_2)]/rk = 0$

$$A_1^* = \frac{1}{rv}\left[\sum_i Q_{i.}^2 + \sum_j Q_{.j}^2 - \frac{2Q_{..}^2}{v}\right] = 321.29$$

$$A_2^* = \frac{1}{r}\left[\sum_i \sum_j Q_{ij}^2 - \frac{1}{v}\sum_i Q_{i.}^2 - \frac{1}{v}\sum_j Q_{.j}^2 + \frac{Q_{..}^2}{v^2}\right] = 296.34$$

Treatment S.S. (adjusted)

$$= \frac{Q_{..}^2}{rv^2} + \sum_{i=1}^{2} \frac{A_1^*}{1 - \mu_1^*} + \sum_s \frac{Q_{ts}^2}{b\beta_s}$$

$$= 2492.40$$

Total S.S. = 4514.00

Block S.S. (unadjusted) = 1656.00

iv) Analysis of variance

ANOVA Table

Source	d.f.	S.S.	M.S.	F
Blocks (unadjusted)	5	1656.00	331.20	
Treatments (adjusted)	9	1492.40	276.93	6.82**
Error	9	365.60	40.62	
Total	23	4514.00		

** indicates significance at 1 per cent level

From the above table it can be inferred that the treatment effects differ significantly.

v) Variance of elementary contrasts

Variance of elementary contrasts are

$$V\left(\hat{t}_i - \hat{t}_j\right) = \frac{2}{rv}\left[\frac{1}{1-\mu_1^*} + \frac{v-1}{1-\mu_2^*}\right] s_e^2$$

$= 48.74 \text{ kg}^2$ when i and j are first associates

$$= \frac{2}{rv}\left[\frac{2}{1-\mu_1^*} + \frac{v-2}{1-\mu_2^*}\right] s_e^2 = 56.87 \text{ kg}^2$$

when i and j are second associates

$$= \left[\frac{1}{b\beta_s} + \frac{1}{rv^2} + \left(\frac{(v-1)^2}{1-\mu_1^*} + \frac{2(v-1)}{1-\mu_2^*} + 1\right)\right] s_e^2$$

$= 32.50 \text{ kg}^2$

one is supplementary and the other is basic treatment.

17

ANALYSIS OF DIALLEL CROSSES (GRIFFING'S APPROACH)

The name diallel which denotes all possible crosses among a group of inbred lines, was introduced by Schmidt (1919). For a set of p inbred lines, we can have p^2 combinations which can be divided into three groups:

1. The 'p' parental lines themselves.
2. one set of p(p-1) /2 F_1's and
3. the set of p(p-1)/2 reciprocal F_1's

Griffing (1956) gave four methods depending upon the material involved.

Method 1 : Includes parents, one set of F_1 and reciprocals

Method 2 : Includes one set of F_1's and parents

Method 3 : includes one set of F_1's and reciprocal

Method 4 : includes one set of F_1's only.

The analysis of diallel crosses is a combining ability analysis which enables the experimenter into splitting the cross effect into 'gca' effects, 'sca' effects and reciprocal effects.

The general combining ability (gca) is the over all performance of lines involved in a set of crosses and any deviation from the average performance is termed as specific combining ability (sca). Here using a suitable statistical model the component variance due to general and specific combining ability are estimated which in turn are translated in to genetical component like σ_A^2 and σ_D^2 under certain assumption. Method of analysis for combining ability considering Eisenhart model 1 (fixed effect) and model II (random effect) has also been given by Griffing.

In each method two steps are involved in the analysis of data. The first step consists of analysis of data for testing the null hypothesis that there are no genotypic differences among the F_1's parents and the reciprocals. When only significant differences are there, then only there is a need for the combining ability analysis.

Method I : (F_1's + parents + reciprocals) (fixed effect model)

Here n parents, (n) (n-1) / 2 F_1's and n(n-1) /2 reciprocal F_1's are laid out usually in randomized block design to find out the difference among genotypes. The model for RBD will be

$$y_{ijk} = \mu + c_{ij} + b_k + e_{ijk} \qquad ...(1)$$

$k = 1, 2 \ldots\ldots r$

$i,j = 1, 2 \ldots\ldots n$

Where μ = general mean

$c_{ij} = (i,j)^{th}$ genotypic mean

b_k is the K^{th} block effect and

e_{ijk} is the random error assumed to be normally and independently distributed with mean 0 and constant variance σ^2.

$$y_{ij} = \sum_{k=1}^{r} y_{ijk} = (i \times j)^{th} \text{ cross totals over all replications}$$

$$y_{..k} = \sum_{i,j=1}^{n} y_{ijk} = k^{th} \text{ block total containing all genotypes}$$

$$y_{...} = \sum_{i,j,k} y_{ijk} = \text{grand total.}$$

Analysis for RBD is done to see if the genotypic differences are found to be significant then combining ability analysis is done. The c_{ij} In model (1) is further splitted as under :

$$c_{ij} = g_i + g_j + s_{ij} + r_{ij}$$

Where g_i is the gca effect of i^{th} line

g_j is the gca effect of j^{th} line

sij is the sca effect of i^{th} and j^{th} line

and r_{ij} is the reciprocal effect of i^{th} and j^{th} lines. Now the full model

becomes;

$$y_{ijk} = \mu + g_i + g_j + s_{ij} + r_{ij} + b_k + e_{ijk}$$

$$i,j = 1, 2 \ldots \ldots n$$

$$k = 1, 2, \ldots r$$

Estimation of combining ability effects and the computation of various sum of square are illustrated with the help of an example :

Example :

Let us consider that 6 parents were involved in full diallel producing 36 crosses laid out in RBD using four replications. The data on grain yield (gms) / plant is given

Crosses	R_1	R_2	R_3	R_4
1 x 1	104.86	84.32	76.92	76.48
1 x 2	88.66	105.04	80.80	73.54
1 x 3	109.76	78.22	74.52	99.52
1 x 4	128.10.	123.84	92.56	115.28
1 x 5	128.36	119.84	103.24	129.72
1 x 6	74.40	70.86	60.94	68.00
2 x 1	88.70	69.10	76.80	88.16
2 x 2	88.02	106.52	89.52	108.68
2 x 3	110.16	116.26	99.76	120.12
2 x 4	101.26	80.22	82.84	88.36
2 x 5	91.52	113.96	87.26	106.98
2 x 6	59.06	65.52	81.62	86.76
3 x 1	75.28	124.74	94.56	114.34
3 x 2	112.48	92.76	90.62	122.36
3 x 3	77.94	71.34	77.52	69.48
3 x 4	114.44	119.96	84.76	86.42
3 x 5	96.88	100.86	86.88	92.52
3 x 6	109.86	98.16	93.26	102.26
4 x 1	124.26	132.48	114.38	105.34

4 x 2	92.18	82.16	81.66	101.24
4 x 3	98.08	90.94	96.2	125.48
4 x 4	80.82	106.54	83.28	95.92
4 x 5	86.20	76.36	79.06	99.52
4 x 6	103.14	109.66	90.98	119.4
5 x 1	109.74	99.56	110.18	125.68
5 x 2	109.94	117.52	95.56	88.54
5 x 3	89.56	94.56	83.66	85.28
5 x 4	80.96	71.98	91.34	89.28
5 x 5	59.96	52.48	52.98	50.98
5 x 6	98.46	73.10	89.18	75.86
6 x 1	72.92	76.28	61.66	64.48
6 x 2	58.56	86.72	65.26	74.64
6 x 3	104.18	100 24	85.12	108.76
6 x 4	109.44	97.74	121.1	106.38
6 x 5	81.58	95.52	84.48	90.28
6 x 6	96.44	98.82	99.14	107.16

Step I : Do the analysis as in case of RBD and prepare the following ANOVA Table.

Table-1 : ANOVA tabel method-I

Source of variation	d.f.	sum of square	Mean Sqaure	F calculated
Replication	3	2029.718		
Treatment	35	32,776.443	936.470	7.942*
Error	105	12381.699	117.921	
Total	**143**	**47,187.860**		

Here the genotypic sum of squares comes out to be significant so we proceed further for the combining ability analysis

Step 2 : Combining ability analysis :

For combining ability analysis the mean of all the crosses over replications are found and are presented in table 2. The total variability in the population may therefore, be partitioned into components like variance due to general combining ability, specific combining ability reciprocal and error. Using replication mean as given in table 2 (one observation per genotype the various sum of squares are obtained as follows :

Table-2 : Mean data of full diallel (Method I)

Parents	1	2	3	4	5	6	Total Mean
1	85.645	87.010	90.505	114.945	120.290	68.550	566.945
2	80.690	98.260	111.575	88.170	99.930	73.240	551.865
3	102.230	104.555	74.070	101.395	94.285	110.885	577.420
4	119.115	89.310	102.675	91.640	85.285	105.795	593.820
5	111.290	102.890	88.265	83.390	54.100	84.150	824.085
6	68.835	71.295	99.575	108.665	87.965	100.390	536.725
Total	567.805	553.320	566.665	588.205	541.855	533.01	3350.86
Total ($y_{i.} + y_{.j}$)	1134.75	1105.185	1144.085	1182.025	1065.940	1069.755	6701.72

Computation of g.c.a. effects

$$\hat{g}_i = \frac{1}{2n}(y_{i.} + y_{.j}) - \frac{1}{n^2} y_{..} \quad \forall\, i = 1,2,\ldots\ldots 6$$

$$\hat{g}_1 = \frac{1}{2 \times 6}(566.945 + 567.805) - \frac{1}{6^2} \times 3350.86$$

$$= 94.5625 - 93.0794 = 1.4831$$

similarly

$$\hat{g}_2 = -0.981$$

$$\hat{g}_3 = 2.261$$

$$\hat{g}_4 = 5.423$$

$$\hat{g}_5 = -4.251$$

$$\hat{g}_6 = -3.935$$

As a check $\Sigma \hat{g}_i = 0$

gca values for the parents are given as the diagonal values in table-3.

Step 3 : Specific combining ability (SCA) Effects

$$s_{ij} = \frac{1}{2}(y_{ij} + y_{ji}) - \frac{1}{2n}(y_{i\cdot} + y_{\cdot i} + y_{j\cdot} + y_{\cdot j}) + \frac{1}{n^2} y...$$

$$s_{12} = \frac{1}{2}(y_{12} + y_{21}) - \frac{1}{2 \times 6}(y_{1\cdot} + y_{\cdot 1} + y_{2\cdot} + y_{\cdot 2}) + \frac{1}{6^2} y...$$

$$= \frac{1}{2}(87.010 + 80.690) - \frac{1}{12}(1134.75 + 1105.185) + \frac{1}{36}(3350.86)$$

$$= 83.850 - 186.661 + 93.079 = \mathbf{-9.732}$$

Similarly other sca effect can be calculated and are presented in the upper half of the table-3.

Step 4 : Reciprocal Effects

The reciprocal effects can be obtained

$$\hat{r}_{ij} = \frac{1}{2}(y_{ij} - y_{ji})$$

When $y_{ij} = y_{ji}$ there will be no reciprocal effect

$$\hat{r}_{21} = \frac{1}{2}(Y_{21} - Y_{12}) = \frac{1}{2}(80.690 - 87.01.) = -3.160$$

All other reciprocal effects have been similarly calculated and are presented in the lower half of table-3.

Table-3 : gca (diagonal) and sca (above diagonal) and reciprocal (below diagonal) effects (method I)

	P(1)	P(2)	P(3)	P(4)	P(5)	P(6)
P(1)	1.483	-9.732	-0.456	17.045	25.479	-21.935
P(2)	-3.160	-.981	13.705	-8.781	13.562	-15.896
P(3)	5.862	-3.510	2.261	1.272	0.186	8.824
P(4)	2.085	0.570	0.640	5.423	-9.913	12.663
P(5)	-4.500	1.480	-3.010	-0.947	-4.251	1.164
P(6)	0.142	-0.973	-0.655	1.435	1.908	-3.935

Step : 5

$$\text{S.S. due to GCA} = \frac{1}{2n}\Sigma\left(y_{i\cdot}+y_{\cdot j}\right)^2 - \frac{2}{n^2}y^2_{\cdot\cdot}$$

$$= \frac{1}{2\times 6}\Big[(566.945 + 567.805)^2 + (551.865 + 553.320)^2$$

$$+.....+(536.725 + 533.01)^2\Big] - \frac{2}{6\times 6}(3350.86)^2$$

$$= 855.071$$

Step 6 :

$$\text{SS due to SCA} = \frac{1}{2}\Sigma\Sigma y_{ij}\left(y_{ij} + y_{ji}\right) - \frac{1}{2n}\Sigma\left(y_{i\cdot}+y_{\cdot j}\right)^2 + \frac{1}{n^2}y^2_{\cdot\cdot}$$

$$= \frac{1}{2}\,[85.645\,(85.645 + 85.645)$$

$$+ 87.010\,(87.010 + 80.690) + \ldots\ldots\ldots\ldots\ldots$$

$$+ 84.150\,(84.150 + 87.965) + \ldots\ldots\ldots\ldots\ldots$$

$$+ 100.390\,(100.390 + 100.390)$$

$$- 1/2x6\,[(566.945 + 567.805)^2 + \ldots\ldots\ldots\ldots$$

$$+ (536.725 + 533.01)^2] + 1/6x6\,(3350.86)^2$$

$$= 7136.6395$$

Step 7 :

$$\text{SS due to reciprocals} = \frac{1}{2}\Sigma\Sigma\left(y_{ij} - y_{ji}\right)^2 \qquad i \neq j$$

$$= \frac{1}{2}\Big[(87.010 - 80.690)^2 + (90.505 - 102.230)^2$$

$$+\ldots\ldots\ldots\ldots\ldots+(84.150 - 87.965)^2\Big]$$

$$= 202.495$$

As a check; treatment SS = r (SSgca + SSsca + SSrec)

SS due to error need not to be calculated. The mean sum of square due to error as obtained in table-2 may be used for this purpose. This value however, be divided by the number of replications (r=4) before using this value for testing the null hypothesis, MS(error) / r is in fact the mean error variance as MSg and MSs have been calculated from the mean data.

The mean error variance required for the F test :

$$MS'(error) = \frac{MS(error)}{n} = \frac{117.921}{4} = 29.480$$

Table 4. Anova for Combing Ability (Method I)

Source	d.f.	SS	MS	F
gca	5	855.07	171.014(Mg)	5.801**
sca	15	7136.6395	475.775(Ms)	16.139**
reciprocal	15	202.495	13.499(Mr)	.4586
error	105		29.48(Me')	

Variance, standard errors and critical differences

$$Var(g_i) = \frac{n-1}{2n^2}\sigma^2 = \frac{6-1}{2\times 36}(29.48) = 2.047$$

$$Var(s_{ii}) = \frac{(n-1)^2}{n^2}\sigma^2 = \frac{25}{36}(29.48) = 20.472$$

$$Var(s_{ij}) = \frac{1}{2n^2}\left(n^2 - 2n + 2\right)\sigma^2 = \frac{1}{2\times 36}(36 - 12 + 2)\times 29.48$$

$$= 10.466$$

$$Var(r_{ij}) = \frac{1}{2}\sigma^2 = \frac{29.48}{2} = 14.74$$

$$Var(g_i - g_j) = \frac{1}{n}\sigma^2 = \frac{29.48}{6} = 4.913$$

$$Var(s_{ij} - s_{ji}) = \frac{2(n-2)}{n}\sigma^2 = \frac{2\times 4}{6}\times 29.48 = 39.307$$

$$Var(s_{ij} - s_{ik}) = \frac{n-1}{n}\sigma^2 = \frac{5}{6} \times 29.48 = 24.567$$

$$Var(s_{ij} - s_{kl}) = \frac{n-2}{n}\sigma^2 = \frac{4}{6} \times 29.48 = 19.653$$

$$Var(r_{ij} - r_{kl}) = \sigma^2 = 29.48$$

These values are used for calculating the critical difference for making comparison between different effects. Now critical difference can be calculated as :

$$C.D. = SEd \times t_{(\alpha/2,\ \text{error d.f})}$$

It we want to compare gca effect of parent 3 and 5 the CD is :

$$CD = \sqrt{Var(g_i - g_j)} \times t_{(.05,105)} = \sqrt{4.913} \times 1.96 = 2.047 \times 1.96 = 4.013$$

Which can be used to compare the mean difference of gca effect of parent 3 and 5. The gca of parent 3 is 2.26 and parent 5 is – 4.251. The difference in 6.511 which is greater than C.D. value. Hence two gca effects differ significantly from each other. Similarly other effects can be compared.

Model 2 (Random Effect Model)

In this case the model will remain the same as in the fixed effect model except that all the components will be assumed random except μ, *i.e.* g_i, g_j, s_{ij}, r_{ij} will be assumed to be normally distributed with mean 0 and variance σ_g^2, σ_s^2 and σ_e^2 respectively, $e_{ijk's}$ are assumed to be normally and independently distributed with mean 0 and constant variance σ^2. The method of computing the various sum of squares will be the same as in the fixed effect model, but the EMS values will be different and instead of estimating the effects corresponding variances are estimated. Here the hypothesis to be tested will be :

$$\sigma_g^2 = 0\ ,\ \sigma_s^2 = 0\ ,\ \sigma_r^2 = 0$$

Genetic component

$$\sigma_g^2 = \frac{1}{2n}\left[M_g - \frac{M_e' + n(n-1)M_s}{n^2 - n + 1}\right]$$

$$= \frac{1}{2 \times 6}\left[171.014 - \frac{29.48 + 6 \times 5 \times 475.775}{36 - 6 + 1}\right]$$

$$= \frac{1}{12}[171.014 - 461.370] = -24.197$$

$$\sigma_s^2 = \frac{n^2}{2(n^2 - n + 1)}(M_s - M'_e) = \frac{36}{2 \times 31}(475.775 - 29.48)$$

$$= 259.139$$

$$\sigma_r^2 = \frac{1}{2}(M_r - M'_e) = \frac{1}{2}(13.49 - 29.48) = -7.995$$

$$\sigma_e^2 = 29.48$$

$\sigma_g^2, \sigma_s^2, \sigma_r^2$ and σ_e^2 stands for estimate

Here $\sigma_g^2 = \frac{1}{2}\sigma_A^2$

and $\sigma_s^2 = \sigma_D^2$

$$\sigma_A^2 = 2 \times \sigma_g^2 = 2 \times -24.197 = -48.394$$

$$\sigma_D^2 = 259.139$$

Method 2 (Parents and one set of F_1's) fixed effect model

Here the model :

$$y_{ijk} = \mu + g_i + g_j + s_{ij} + b_k + e_{ijk} \qquad k = 1,2.....r$$

$$r \leq j = 1, 2, 3.....n$$

Considering n = 6 there will be 21 crosses and 6 parents. The data from example I has been taken for illustration and is given in Table-5.

Table 5 : Data an grain yield (gm) plot

	R_1	R_2	R_3	R_4
1 x 1	104.86	84.32	76.92	76.48
1 x 2	88.66	105.04	80.80	73.54
1 x 3	109.76	78.22	74.52	99.52
1 x 4	128.10.	123.84	92.56	115.28
1 x 5	128.36	119.84	103.24	129.72
1 x 6	74.40	70.86	60.94	68.00
2 x 2	88.02	106.52	89.52	108.68
2 x 3	110.16	116.26	99.76	120.12
2 x 4	101.26	80.22	82.84	88.36
2 x 5	91.52	113.96	87.26	106.98
2 x 6	59.06	65.52	81.62	86.76
3 x 3	77.94	71.34	77.52	69.48
3 x 4	114.44	119.96	84.76	86.42
3 x 5	96.88	100.86	86.88	92.52
3 x 6	109.86	98.16	93.26	102.26
4 x 4	80.32	106.54	83.28	95.92
4 x 5	86.20	76.36	79.06	99.52
4 x 6	103.14	109.66	90.98	119.40
5 x 5	59.96	52.48	52.98	50.98
5 x 6	98.46	73.10	89.18	75.86
6 x 6	96.44	98.82	99.14	107.16

Example :

Step 1 : The various sum of squares are calculated as in RBD and the ANOVA table will be :

Table 6 : ANOVA for Method 2

Source	d.f.	sum of square	Mean Square	F-Calculated
Replication	3	1723.821	-	-
Treatment	20	20375.372	1018.769	9.565*
Error	60	6390.778	106.513	-
Total	**83**	**28489.971**	-	-

Here the genotypic sum of square comes out to be significant so we proceed further for the combining ability analysis

2. Combining ability analysis

The data given in table-5 is rearranged by taking their mean. Thus each genotype is represented by one observation only as is given in table-7 :

Table 7 : Replication Mean in Method 2

Parents	P_1	P_2	P_3	P_4	P_5	P_6	Total ($Y_{i.}$)	$Y_{i.} - Y_{ii}$	$Y_{i.} + y_{ii}$
P_1	85.645	87.010	90.505	114.945	120.290	68.550	566.945	481.30	652.590
P_2	87.010	98.260	111.575	88.170	99.930	73.240	558.185	372.915	656.445
P_3	90.505	111.575	74.070	101.395	94.285	100.885	572.715	296.565	646.785
P_4	114.945	88.170	101.395	91.640	85.285	105.795	587.23	191.08	678.87
P_5	120.290	99.930	94.285	85.285	54.120	84.150	538.04	84.150	592.14
P_6	68.550	73.240	100.885	105.795	84.150	100.390	533.01	-	633.40
Total	88.645	185.270	276.15	396.150	453.89	533.010	1930.115		3860.23

Step-2 : Estimating of GCA effects

The general combining ability effects are defined as follows :

$$g_i = \frac{1}{(n+2)}\left[(y_{i.}+y_{ii}) - \frac{2}{n}y_{..}\right]$$

$$g_1 = \frac{1}{8}\left[(y_{i.} + y_{ii}) - \frac{2}{6}y_{..}\right]$$

$$g_1 = \frac{1}{8}\left[652.590 - \frac{2}{6} \times 1930.115\right] = \mathbf{+1.152}$$

Similarly

$$g_2 = \frac{1}{6}\left[(y_{2.} + y_{22}) - \frac{2}{6}y_{..}\right]$$

$$g_2 = \frac{1}{6}\left[656.445 - \frac{2}{6} \times 1930.115\right] = \mathbf{+1.634}$$

$g_3 = 0.427$, $g_4 = 4.437$, $g_5 = -6.404$, $g_6 = -1.246$

gca effects are given in the diagonal of table-8

Step-3 : Specific combining ability (sca) effects

$$s_{ij} = y_{ij} - \frac{1}{(n+2)}(y_{i.}+y_{ii} + y_{.j}+y_{jj}) + \frac{2}{(n+1)(n+2)}y_{..}$$

$$s_{12} = y_{12} - \frac{1}{8}(y_{1.}+y_{11} + y_{.2}+y_{22}) + \frac{2}{7 \times 8}y_{..}$$

$$= 87.010 - \frac{1}{8}(652.590 + 656.445) + \frac{2}{56} \times 1930.115$$

$$= 87.010 - 163.629 + 62.932 = \mathbf{-7.687}$$

Similarly sca effects for all other crosses are estimated and are given in upper half of the table-8.

Step 4 : Estimation of sum of square

SS due to gca

$$= \frac{1}{(n+2)}\left[\Sigma(y_{i\cdot}+y_{ii})^2 - \frac{4}{n}y_{\cdot\cdot}^2\right]$$

$$= \frac{1}{8}\left[(652.590)^2 + (656.445)^2 + (646.785)^2 + (678.87)^2\right.$$

$$\left. + (59214)^2 + (633.40)^2 - \frac{4}{6}\times(1930.115)^2\right]$$

$$= \frac{1}{8}[2487814.399 - 2483562.609]$$

$$= \frac{4251.7902}{8} = \mathbf{531.473}$$

Table-8 : gca (diagonal) and sca (above diagonal) effects for Method-2

	P_1	P_2	P_3	P_4	P_5	P_6
P_1	1.152	-7.687	-2.984	17.445	33.631	-23.266
P_2	-	1.634	17.604	-9.812	12.790	-19.058
P_3	-	-	0.427	4.621	8.352	9.795
P_4	-	-	-	+ 4.437	-4.659	10.694
P_5	-	-	-	-	-6.404	-0.110
P_6	-	-	-	-	-	-1.246

Step 5 : SS due to sca

$$= \sum_{i\le j}\Sigma y_{ij}^2 - \frac{1}{(n+2)}\Sigma(y_{i\cdot}+y_{ii})^2 + \frac{2}{(n+1)(n+2)}y^2{}_{\cdot\cdot}$$

$$= (85.645)^2 + (87.010)^2 \ldots + (54.10)^2 + (84.15)^2 + (100.39)^2$$

$$- \frac{1}{8}\left[(652.590)^2 + \ldots + (633.40)^2\right] + \frac{2}{7\times 8}(1930.115)^2$$

$$= 182491.1032 - \frac{1}{8}(2487814.399) + 118264.886$$

$$= 182491.1032 - 310976.7999 + 133047.99 = \mathbf{4562.2933}$$

Check : Treat SS = r (SS due to gca + SS due to sca).

SS due to error is obtained from the table by dividing the error SS by 4. Thus SS due to error is 1597.6945.

The analysis of variance for combining ability analysis in Method 2 is given in Table 9.

Table-9 : ANOVA for combining ability analysis (Method 2)

Source	d.f	SS	MS	FF_{cal}
gca	5	531.473	106.2946 (Mg)	3.99
sca	15	4562.2933	304.152 (Ms)	11.42*
Error	60	1597.6945	26.6282(M_e')	

Variance, standard errors and critical difference

$$Var(g_i) = \frac{(n-1)}{n(n+2)}\sigma^2 = \frac{5}{6\times 8} \times 26.6282 = 2.774$$

$$Var(s_{ij}) = \frac{(n^2+n+2)}{(n+1)(n+2)}\sigma^2 = \frac{(36+6+2)}{7\times 8} \times 26.6282 = 20.922$$

$$Var(g_i - g_j) = \frac{2\sigma^2}{(n+2)} = \frac{2}{8} \times 26.6282 = 6.657$$

$$Var(s_{ij}) = \frac{n(n-1)}{(n+1)(n+2)}\sigma^2 = \frac{6\times 5}{7\times 8} \times 26.6282 = 14.265$$

$$Var(s_{ij} - s_{jj}) = \frac{2(n-2)}{(n+2)}\sigma^2 = \frac{2\times 4}{8} \times 26.6282 = 26.6282$$

$$Var(s_{ij} - s_{ik}) = \frac{2(n+1)\sigma^2}{(n+2)} = \frac{2\times 7}{8} \times 26.6282 = 46.5993$$

$$Var(s_{ij} - s_{kl}) = \frac{2n}{(n+2)}\sigma^2 = \frac{2\times 6}{8} \times 26.6282 = 39.9423$$

The various standard errors can be calculated by taking the square root of the various variance and critical difference can be found in the usual way.

Method 2 (Random Effect Model)

Estimation of component variance :

$$\sigma_g^2 = \frac{1}{(n+2)}(M_g - M_s)$$

$$= \frac{1}{8}(106.2946 - 304.152) = \mathbf{-24.732}$$

$\sigma_s^2 = M_s - M_e' = 304.152 - 26.6282 = 277.5238$

These component can be translated into genetic component using the following equation

$$\sigma_g^2 = \frac{1}{2}\sigma_A^2, \quad \sigma_A^2 = 2\sigma_g^2 = 2 \times -24.732 \quad = -49.464$$

$$\sigma_s^2 = \sigma_D^2 \qquad \sigma_D^2 = 277.5238$$

Method 3 (one set of F_1's and reciprocals)

Fixed Effect Model :

In this method one set of F_1's and reciprocal crosses are used. The total entries are n (n-1) F_1's and n parents. The data given in table-1 after ignoring the values of diagonal *i.e.* the parent value has been used for illustrating the method of analysis which is given in Table-10.

Table-10 : grain yield (gms) / Plant

	R1	R2	R3	R4
1 x 2	88.66	105.04	80.80	73.54
1 x 3	109.76	78.22	74.52	99.52
1 x 4	128.10	123.84	92.56	115.28
1 x 5	128.36	119.84	103.24	129.72
1 x 6	74.40	70.86	60.94	68.00
2 x 1	88.70	69.10	76.80	88.16

2 x 3	110.16	116.26	99.76	120.12
2 x 4	101.26	80.22	82.84	88.36
2 x 5	91.52	113.96	87.26	106.98
2 x 6	59.06	65.52	81.62	86.76
3 x 1	75.28	124.74	94.56	114.34
3 x 2	112.48	92.76	90.62	122.36
3 x 4	114.44	119.96	84.76	86.42
3 x 5	96.88	100.86	86.88	92.52
3 x 6	109.86	98.16	93.26	102.26
4 x 1	124.26	132.48	114.38	105.34
4 x 2	92.18	82.16	81.66	101.24
4 x 3	98.08	90.94	96.20	125.48
4 x 5	86.20	76.36	79.06	99.52
4 x 6	103.14	109.66	90.98	119.4
5 x 1	109.74	99.56	110.18	125.68
5 x 2	109.94	117.52	95.56	88.54
5 x 3	89.56	94.56	83.66	85.28
5 x 4	80.96	71.98	91.34	89.28
5 x 6	98.46	73.10	89.18	75.86
6 x 1	72.92	76.28	61.66	64.48
6 x 2	58.56	86.72	65.26	74.64
6 x 3	104.18	120.24	85.12	108.76
6 x 4	109.44	97.74	121.1	106.38
6 x 5	81.58	95.52	84.48	90.28

1. Testing the significance of genotypic difference :

The various sum of squares are calculated as in RBD. The analysis of variance (Method 3) is given in Table-11.

Table-11 : Analysis of Variance (Method 3)

Source	d.f.	SS	M.S.	F
Replication	3	1980.881		
Treatment	29	24308.736	838.232	6.60
Error	87	10950.440	125.867	
Total	119	37,240.056		

2. Combining Ability Analysis :

As done earlier the combining ability analysis is done over mean value ignoring the diagonal values as given in Table-12.

Table-12 : Mean Data

	P(1)	P(2)	P(3)	P(4)	P(5)	P(6)	Total Yi
P(1)	-	87.010	90.505	114.945	120.290	68.550	481.300
P(2)	80.690	-	111.575	88.170	99.93	73.240	453.605
P(3)	102.230	104.555	-	101.395	94.285	100.885	503.350
P(4)	119.115	89.310	102.675	-	85.285	105.795	502.180
P(5)	111.290	102.890	88.265	83.390	-	84.150	469.985
P(6)	68.835	71.295	99.575	108.665	87.965	-	436.335
Total y.j	482.16	455.06	492.595	496.565	487.755	432.62	
(yi.+y.j)	963.46	908.665	995.945	998.745	957.74	868.955	2846.755

Estimation of gca effects

$$g_i = \frac{1}{2n(n-2)}\left[n(y_{i\cdot}+y_{\cdot i}) - 2y_{\cdot\cdot}\right]$$

$$g_1 = \frac{1}{2\times 6\times 4}\left[6\times 963.46 - 2\times 2846.755\right]$$

$$g_1 = \frac{1}{48}\left[5780.76 - 5693.51\right] = 1.817$$

Similarly

$g_2 = -5.032$

$g_3 = 5.878$

$g_4 = 6.228$

$g_5 = 1.103$

$g_6 = -9.995$

and are shown in the diagonal values of table-13.

Estimation of sca effects

$$s_{ij} = \frac{1}{2}(y_{ij} + y_{ji}) - \frac{1}{2(n-2)}(y_{i\cdot} + y_{\cdot i} + y_{j\cdot} + y_{\cdot j}) + \frac{1}{(n-1)(n-2)} y_{\cdot\cdot}$$

$$s_{12} = \frac{1}{2}(y_{12} + y_{21}) - \frac{1}{2\times 4}(y_{1\cdot} + y_{\cdot 1} + y_{2\cdot} + y_{\cdot 2}) + \frac{1}{5\times 4} y_{\cdot\cdot}$$

$$s_{12} = \frac{1}{2}(87.010 + 80.690) - \frac{1}{8}(963.46 + 908.665)$$

$$+ \frac{1}{20} \times 2846.755$$

$= 83.85 - 234.0156 + 142.3377 =$ **-7.828**

Similarly other sca effects for the crosses are calculated and are given in the upper half of Table-13.

$$r_{ij} = \frac{1}{2}(y_{ij} - y_{ji})$$

$$r_{21} = \frac{1}{2}(80.690 - 87.010) = -3.16$$

The other reciprocal effects are given in the lower half of the table-13.

Step 4: S.S. due to gca :

$$= \frac{1}{2(n-2)} \Sigma(y_{i\cdot} + y_{\cdot i})^2 - \frac{2}{n(n-2)} y^2_{\cdot\cdot}$$

$$= \frac{1}{2 \times 4}\left[(963.46)^2 + (908.665)^2 + \ldots\ldots + (868.955)^2\right]$$

$$- \frac{2}{6 \times 4}(2846.755)^2$$

$$= 676959.2464 - 6575334.5025 = \mathbf{1624.7739}$$

Step 5 : S.S. due to sca :

$$= \frac{1}{2}\Sigma\Sigma\left(y_{ij} + y_{ji}\right)^2 - \frac{1}{2(n-2)}\Sigma\left(y_{i\cdot} + y_{\cdot j}\right)^2 + \frac{1}{(n-1)(n-2)}y^2_{\cdot\cdot}$$

$$= \frac{1}{2}\left[(87.010 + 80.696)^2 + (90.505 + 102.330)^2 + \ldots\ldots +\right.$$

$$\left.(84.150 + 87.965)^2\right] - \frac{1}{2 \times 4}\left[(963.46)^2 + (908.665)^2 + \ldots\ldots +\right.$$

$$\left.(868.955)^2\right] + \frac{1}{5 \times 4}(2846.755)^2$$

$$= \frac{1}{2} \times 552016.9338 - 676959.2464 + 405200.7015$$

$$= 276008.4669 - 676959.2464 + 405200.7015$$

$$= \mathbf{4249.922}$$

$$\text{SS due to reciprocal} = \frac{1}{2}\Sigma\Sigma\left(y_{ij} - y_{ji}\right)^2$$

$$= \frac{1}{2}\left[(87.010 - 80.690)^2 + (90.505 - 102.230)^2 + \ldots\ldots\right.$$

$$\left.(84.150 - 87.965)^2\right] = 202.4881$$

Check :

Treat SS = r (SS gca + SSsca + SS reciprocal)

$$\text{SS due to error } (M_e') = \frac{\text{S.S due to error (table 11)}}{4}$$

$$= \frac{10950.44}{4} = 2737.61$$

Table 13 : gca (diagonal) sca (above diagonal) and reciprocal below diagonal effects

	P(1)	P(2)	P(3)	P(4)	P(5)	P(6)
P(1)	1.818	-7.828	-6.220	14.092	17.978	-18.022
P(2)	-3.160	-5.032	12.326	-7.349	10.447	-7.597
P(3)	5.862	-3.510	5.878	-4.964	-10.598	9.455
P(4)	2.085	0.570	0.640	6.228	-17.885	16.105
P(5)	-4.500	1.480	-3.010	-0.947	1.103	0.058
P(6)	0.142	-0.973	-0.655	1.435	1.908	-9.995

The analysis of variance is given in table-14.

Table 14 : ANOVA for combining ability analysis (Method 3)

Source	d.f.	SS	MS	F
gca	5	1624.7739	324.9548 (M_g)	10.326*
sca	9	4249.922	472.2135 (M_s)	15.007*
Reciprocal	15	202.4881	13.4992 (M_r)	
Error	87	-	31.4667 (M_e')	

Estimation of variance, standard error and the critical differences

$$Var(g_i) = \frac{(n-1)}{2n(n-2)}\sigma^2 = \frac{5}{2\times 6\times 4} \times 31.4667 = 3.2778$$

$$Var(s_{ij}) = \frac{(n-3)\sigma^2}{2(n-1)} = \frac{3}{2\times 5} \times 31.4667 = 9.4400$$

$$Var(r_{ij}) = \frac{\sigma^2}{2} = \frac{31.4667}{2} = 15.7333$$

$$Var(g_i - g_j) = \frac{\sigma^2}{(n-2)} = \frac{31.4667}{4} = 7.8666$$

$$Var(s_{ij} - s_{ik}) = \frac{(n-3)}{n-2}\sigma^2 = \frac{3}{4} \times 31.4667 = 23.6000$$

$$Var(s_{ij} - s_{ki}) = \frac{(n-4)}{(n-2)}\sigma^2 = \frac{2}{4} \times 31.4667 = 15.7333$$

Now various standard errors and C.D. values can be calculated.

Method 3 : Random effect Model :

Estimation of component of variance and their genetic interpretations

$$\sigma_g^2 = \frac{1}{2(n-2)}(M_g - M_s) = \frac{1}{2 \times 4}(324.9548 - 472.2135)$$

= - 18.4073

$$\alpha_s^2 = \frac{1}{2}(M_s - M_e') = \frac{1}{2}(472.2135 - 31.4667) = 220.3734$$

$$\alpha_r^2 = \frac{1}{2}(M_s - M_e') = \frac{1}{2}(13.4992 - 31.4667) = -8.9837$$

$$\alpha^2 = M_e' = 31.4667$$

where

$$\sigma_g^2 = \frac{1}{2}\sigma_A^2 \quad \text{and} \quad \sigma_s^2 = \sigma_D^2$$

$$\sigma_A^2 = 2x - 18.4073 = -36.8146$$

$$\sigma_D^2 = 220.3734 = 220.3734$$

Method 4 (One set of F_1's only)

Fixed effect Model

The experimental material consists of F_1s only, the total entry being (n(n-1)/ 2 (*i.e.* 15 in case the number of parents is 6. The data given in table 5 (Method 2) has been used for the analysis after ignoring the parents

The Analysis of variance is given in table-15.

Table 15 : Analysis of Variance (Method 4)

Source of variation	d.f.	Sum of sqaure	Mean Square	Feal
Replication	3	1785.694	-	-
Treatment	14	12179.294	869.950	7.535*
Error	42	4848.807	115.448	
Total	**59**	**18813.796**		

Combining Ability Analysis

Means of data over replications are calculated as in the previous methods for preparing the diallel table required for combining ability analysis :

Table 16 : Mean data for Full diallel (Method 4)

	P(1)	P(2)	P(3)	P(4)	P(5)	P(6)	Total
P(1)	-	87.010	90.505	114.945	120.290	68.550	481.300
P(2)	87.010	-	111.575	88.170	99.930	73.240	459.925
P(3)	90.505	111.575	-	101.395	94.285	100.885	498.645
P(4)	114.945	88.170	101.395	-	85.285	105.795	495.59
P(5)	120.290	99.930	94.285	85.285	-	84.150	483.94
P(6)	68.550	73.240	100.885	105.795	84.150	-	1426.01

Estimation of the gca effects

$$g_i = \frac{1}{n(n-2)}\left[n\, y_{i\cdot} - 2y_{\cdot\cdot}\right]$$

$$g_1 = \frac{1}{6 \times 4}\left[6 \times 481.300 - 2 \times 1426.01\right]$$

$$g_1 = \frac{1}{24}\left[2887.8 - 2852.02\right] = 1.491$$

Similarly

$g_2 = -3.853$
$g_3 = 5.827$
$g_4 = 5.062$
$g_5 = 2.151$

$g_6 = -10.679$

These values are given as diagonal elements in Table 17.

Estimation of sca effects

$$s_{ij} = y_{ij} - \frac{1}{(n-2)}(y_{i\cdot}+y_{\cdot j}) + \frac{2}{(n-1)(n-2)}y_{\cdot\cdot}$$

$$s_{12} = y_{12} - \frac{1}{4}(y_{1\cdot}+y_{\cdot 2}) + \frac{2}{5\times 4}(1426.01)$$

$$= 87.010 - \frac{1}{4}(481.30 + 459.925) + \frac{2}{20}\times 1426.01$$

$$= 87.010 - 235.3062 + 142.601 = \mathbf{-5.695}$$

Similarly the other sca effects can be calculated for the remaining error and are presented in Table-17.

Combining Ability Analysis

Estimation of sum of squares :

$$\text{SS due to gca} = \frac{1}{(n-2)}\Sigma y_{i.}^2 - \frac{4}{n(n-2)}y^2_{\cdot\cdot}$$

$$= \frac{1}{4}\left[(481.30)^2 + (459.925)^2 .....+(432.62)^2\right] - \frac{4}{6\times 4}(1426.01)^2$$

$$= \frac{1}{4}\left[1358794.968\right] - \frac{1}{6}(1426.01)^2$$

$$= 339698.7419 - 338917.42 = \mathbf{781.3218}$$

Table 17 : gcs (diagonal), sca (above diagonal) effects

P(1)	P(2)	P(3)	P(4)	P(5)	P(6)	Total
P(1)	1.491	-5.965	-11.880	13.324	21.581	-17.329
P(2)		-3.857	14.533	-8.108	6.565	-7.295
P(3)			5.827	-4.563	-8.700	10.670
P(4)				5.062	-16.997	16.343
P(5)					2.151	-2.389
P(6)						-10.679

SS due to sca

$$\Sigma\Sigma y_{ij}^2 - \frac{1}{n-2}\Sigma y_{i.}^2 + \frac{2}{(n-1)(n-2)} y^2..$$

$$= (87.01)^2 + (90.505)^2 .....+(84.105)^2 - \frac{1}{4}\Big[(481.30)^2$$

$$+(459.925)^2 +.....+(432.62)^2\Big] + \frac{2}{5\times 4}(1426.01)^2$$

$$= 138611.793 - 339698.7419 + 203350.452$$

$$= \mathbf{2263.5031}$$

Check :

Treat SS = r (SS due to gca + SS due sca)

Table 18 : ANOVA for combining ability Analysis (Method 4)

Source	d.f.	Sum of sqaure	Mean Square	Feal
gcs	5	781.3218	156.2644 (Mg)	5.414
sca	9	2263.5031	251.500 (Ms)	8.714
error	42		M'e = 28.862	

$$\text{SS due to error } (M_e') = \frac{\text{Error mean square}}{4} \text{ (Table 15)}$$

$$= \frac{115.448}{4} = 28.862$$

Estimation of standard Error

$$SE(g_i) = \frac{n-1}{n(n-2)}\sigma^2 = \frac{5}{6\times 4} \times 28.862 = 6.0129$$

$$SE(s_{ij}) = \frac{n-3}{n-1}\sigma^2 = \frac{3}{5} \times 28.862 = 17.3172$$

$$SE(g_i - g_j) = \frac{2}{(n-2)}\sigma^2 = \frac{2}{4} \times 28.862 = 14.431$$

$$SE(s_{ij} - s_{ik}) = \frac{2(n-3)}{(n-2)}\sigma^2 = \frac{2 \times 3}{4} \times 28.862 = 43.293$$

$$SE(s_{ij} - s_{kl}) = \frac{2(n-4)\sigma^2}{(n-2)} = \frac{2 \times 2}{4} \times 28.862 = 28.862$$

The various standard error and CD values can be calculated in the usual way.

Method 4 (Random effect Model)

Estimation of Component of Variance and their genetic interpretation

$$\sigma_g^2 = \frac{1}{(n-2)}(M_g - M_s) = \frac{1}{4}(156.2644 - 251.500)$$

$$= -23.809$$

$$\alpha_s^2 = M_s = M_e' = 251.500 - 28.862 = 222.638$$

Further

$$\sigma_g^2 = \frac{1}{2}\sigma_A^2 \qquad \sigma_A^2 = 2 \times -23.809 = -47.618$$

$$\sigma_s^2 = \sigma_D^2 \qquad \sigma_D^2 = 222.638$$

❑❑❑

18

DESIGNS AND ANALYSIS FOR BIOLOGICAL ASSAYS

A biological assay is an experiment conducted on the living organism with a view to compare the relative efficiency of two (or more) materials like hormones, drugs, insecticides, *etc.* on the basis of the effects that they produce on the living organisms. Here the comparison is made on the basis of two sets of doses, one from the standard preparations and the other from test preparations. The standard preparation is likely to be sample of either the international standard or a more readily available working standard whose potency relative to the international standard has previously been worked out carefully and the test preparation is unknown potency whose relative potency is to be evaluated. If Zs and Zt denote the doses of the standard and test preparations respectively, which produce the same response, then comparison of two test preparations is obtained by calculating the coefficient called the relative potency, which is given by the ratio p = Zs / Zt. Thus the central purpose of a bio essay is to estimate the strength of a test preparation in relation to the standard preparation in terms of the relative potency alongwith a measure of its precision by conducting an appropriate experiment for estimating the doses of S and T.

Every bio assay has three components namely :

i) Stimuli (materials) which are to the compared

ii) Living organisms (subjects) to which different stimuli are given.

iii) A measure of characteristic response produced by the stimuli on the organism.

For practical examples refer to Finney (1978, chapter 4 and 5)

A bio assay can be either direct or indirect. For direct assays the doses of standard and test preparations are just sufficient to produce the desired effect, for example death of an animal. Direct assays are used only when both the preparations can be administered in such a way that the minimal amounts needed to produce the specified response can be measured exactly. In indirect assays predetermined doses are administered to the subjects and their responses are recorded. If the two preparation involved in an assay contain the same effective ingrediant and all the other substance, are inert, then the assay is called an analytical dilution assay. Here we will discuss only indirect diluting biological assays.

Metametric Transformation

The available methods for estimation of relative potency are based on the assumption of linear relationship between doses and the response. If the relationship is not linear, either dose or response or both are transformed to make the relationship linear. These transformations are commonly known as linearising or metametric transformation. The transformed dose and response are called dose metameter and response metameter. The usual transformation are of the type (i) log Z = λ, y and (ii) z^{λ} = x,y where x is the dose metameter, y is the response, Z is the dose in the original scale and is a constrant. As a consequence of these transformations, the indirect bio assays have been divided in to categories (a) parallel line assays (b) slope ratio assays.

Case 1: For x = log (dose) as linerising transformation, let $Y = a_s + bx_s$ denote the relation between the response. y and x, where x_s = log z_s and z_s denote dose of standard preparation. If z_t denote dose equipotent of the test preparation then $p = z_s / z_t$ *i.e.* log p = log z_s – log $z_t = x_s - x_t$ as x_s = log p + x_t. We have

$$Y = a_s + b (\log p + x_t) \text{ that is}$$

$$Y = a_t + b_{xt} \text{ where } a_t = a_s + b \log p$$

The relationship of test preparation is also linear like that of a standard preparation. These lines for standard and test preparation have the same slope and hence are parallel. Here $\log p = (a_t - a_s) / b$ with a_t and a_s the intercepts and b is the common slope.

If there are K doses of each of two preprataions and x_s and x_t denote the averages of the dose metameter and $\bar{y}_s$ and $\bar{y}_t$ are the average responses for the preparation then it is known that : -

$$a_s = \bar{y}_s - b\bar{x}_s \quad \text{and} \quad a_t = \bar{y}_t - b\bar{x}_t$$

on substituiting these values in log p = $(a_t - a_s) / b$ we get an estimate R of p as log $R = \bar{x}_s - \bar{x}_t - (\bar{y}_s - \bar{y}_t) / b$

Two lines are parallel when the dose metameter is log (dose). The assays corresponding to this transformation are called parallel line assays.

Here $(\bar{y}_s - \bar{y}_t)$ is a contrast and it is called the preparation contrast. A paralel line assay is said to be symmetrical (SPL) if the number of doses in each of the preparation is same otherwise it is called asymmetrical (APL).

Case II : When the linearising transformation is $x = z^{\lambda}$ the two equations are $y = a + b_s x_s$ and $y = a + b_t + x_t$ for standard and test preparations where $b_t = b_s p^{\lambda}$ and therefore, $p^{\lambda} = b_t/b_s$. The corresponding assays are called slope ratio assays. These lines interact at the response axis. This gives rise to another validity test in addition to the validity test for linear relationship.

Parallel line assays :

Suppose there are K doses for each preparation, and n subjects be allotted to each of these preparations. Let the doses of standard and test preparations be respectively denoted by $s_1, s_2 \ldots\ldots s_k$ and $t_1, t_2 \ldots\ldots\ldots\ldots t_k$. The response of rth subject allotted to p^{-th} (g^{-th}) dose of standard (test) preparation be denotely $y_s p_r$ ($y_t p_r$). The response data can be arranged in the form.

Response data from 2k point Assay

Response	Standard preparation doses				Test preparation doses			
	s_1, s_2			s_k	t_1, t_2			t_k
	y_{s11}	y_{s21}		y_{sk1}	y_{t11}	y_{t21}	.....	y_{tk1}
	y_{s12}	y_{s22}	.....	y_{sk2}	y_{t12}	y_{t22}	.....	y_{tk2}
	.....	.....		.....	.....	.....		.....
	.....	.....		.....	.....	.....		.....
	y_{s1n}	y_{s2n}		y_{skn}	y_{t1n}	y_{t2n}	.....	y_{tkn}
Total	S_1	S_2	.....	S_k	T_1	T_2	.....	T_k

The analysis and the estimation of the relative potency becomes very

much simplified when the doses of each of the preparations are taken in geometric progression as given below.

s, cs, c^2s..... $c^{k-1}s$ and t, ct, c^2t, ... $c^{k-1}t$ where c is a constant while s and t are the initial doses of standard and test preparation.

The doses should be evenly distributed in the range of response in which the dose response relationship was investigated for obtaining the linearing transformations.

$$x_{si} = \log c^i s = \log s + i \log c \quad (i = 0, 1, 2 \ldots\ldots\ldots\ldots k-1)$$

$$x_{ti} = \log c^i t = \log t + i \log c \quad (i = 0, 1, 2 \ldots\ldots\ldots\ldots k-1)$$

Here

$$\bar{x}_s = \log s + \frac{k-1}{2} \log c$$

$$\bar{x}_t = \log t + \frac{k-1}{2} \log c$$

and

$$x_{si} - \bar{x}_s = \left(i - \frac{k-1}{2}\right) \log c \text{ and}$$

$$x_{ti} - \bar{x}_t = \left(i - \frac{k-1}{2}\right) \log c$$

Here $\bar{x}_s$ and $\bar{x}_t$ denotes the averages of two preparations. Now two cases arise when k is odd and when k is even.

Case I : When k is odd we choose the base of the logarithm as c so that log c is 1. The log dose as deviates from their mean can now be written as

Standard preparations :

$$\frac{-k-1}{2}, \frac{-k-3}{2}, \ldots\ldots\ldots -1, 0, 1, \ldots\ldots\ldots \frac{k-1}{2}$$

Test preparation :

$$\frac{-k-1}{2}, \frac{-k-3}{2}, \ldots\ldots\ldots -1, 0, 1, \ldots\ldots\ldots \frac{k-1}{2}$$

Case II: When k is even, the base of log is taken as $\sqrt{c}$ so that log c becomes 2 and hence all the dose deviate [(i-k-1)/2] log c becomes odd integers as shown below :

Standard preparation :

-(k-1), -(k-3),.....-1, 1, 3,.....,(k-1)

Test preparation :

-(k-1) -(k-3),..... -1, 1, 3, ...(k-1)

The regression contrast in each preparation can be obtained by multiplying these deviates by the corresponding dose totals and adding them.

Analysis :

Here first through Analysis of variance technique it is tested (1) the response and dose metameter relationship is linear (2) two lines are perpendicular to each other. If both these validity tests are found to be correct then the relative potency is estimated from :

$$\log R = \bar{x}_s - \bar{x}_t - \frac{\bar{y}_s - \bar{y}_t}{b} \text{ and we have}$$

$$\bar{y}_s - \bar{y}_t = \frac{\sum_i s_i - \sum_i T_i}{kn}$$

For the first part of the analysis, preparation contrast among the dose totals is obtained as:

$$\text{Preparation contrast (Lp)} = -\sum_i s_i + \sum_i T_i$$

when K is odd, the combined regression contrast is given as :

$$(L_1) = \frac{-(k-1)}{2}(S_1 + T_1) - \frac{k-3}{2}(S_2 + T_2).....+\frac{k-1}{2}(S_k + T_k)$$

and when K is even

$$(L_1) = -(k-1)(S_1 + T_1) - (k-3)(S_2 + T_2).....+(k-1)(S_k + T_k)$$

The difference between the two regression contrast of the two preparations is called the parallelism contrast.

Parallelism contrast (L_1)

$$= \frac{-(k-1)}{2}(S_1 - T_1) - \frac{k-3}{2}(S_2 - T_2)....-\frac{(k-1)}{2}(S_k - T_k)$$

when k is odd.

and when k is even

$$L_1' = -(k-1)(S_1 - T_1) - (k-3)(S_2 - T_2) \ldots + (k-1)(S_k - T_k)$$

Here it can be seen that

$$\bar{y}_s - \bar{y}_t = \frac{-L_p}{kn}$$

When k is odd, $b = \frac{6L_1}{kn(k^2-1)}$ and

When k is even, $b = \frac{3L_1}{2kn(k^2-1)}$

The following analysis variance can be written.

Analysis of variance in 2k point assays for validity tests

Source of Variation	d.f.	s.s.	m.s	f
Preparation (Lp)	1	$L_P^2 / 2kn$		
Combined Regression (L_1)	1	L_1^2 / D		
Parallalism (L_1')	1	L_i^2/D	s^2b	s^2b/s^2e
Deviation from regression	2k-1	By Subtraction	s^2d	s^2d/s^2e
Doses	2k-1	$\frac{\Sigma S_i^2 + \Sigma T_i^2}{n} - \frac{\{\Sigma(S_i + T_i)\}^2}{kn}$		
with in doses (error)	2k(n-1)	By subtraction	s^2e	
Total	2kn-1	$\sum_{p_r} y_s^2 p_r + \sum_{q_r} y_t^2 q_r - \frac{\Sigma(S_i + T_i)^2}{kn}$		

Here $D = kn(k^2-1)/6$ when k is odd and

$D = 2kn\,(k^2-1)\,/3$ when k is even.

For testing linearity of regression, the mean square for the deviation from regression is tested by F test. For testing parallelism the "parallelism" components is tested.

If both these tests are not significant, then the relative potency can be calculated

$$\log R \qquad \bar{x}_s - \bar{x}_t - \frac{\bar{y}_s - \bar{y}_t}{b}$$

$$= \log s - \log t + \frac{Lp}{kn}\frac{kn(k^2-1)j}{6L} \qquad \text{when k is odd.}$$

$$R \qquad = \frac{s}{t}\text{antilog}\left[\frac{d(k^2-1)}{6}\frac{LP}{L_1}\right]$$

Here $d = \log_{10}^{C}$

When k is even

$$\log R = \log\frac{s}{t} + \frac{L_p}{kn}\frac{2kn(k^2-1)}{3L_l}j$$

$$R = \frac{s}{t}\text{antilog}\left[\frac{d(k^2-1)}{3}\frac{LP}{L_l}\right]$$

Here we can use CRD or RBD depending upon the availably of homogeneous groups of experimental material. These designs are not applicable when k is greater than 3 or 4. Usually small animals Like rats, cats, guina piga, *etc.* are used as experimental units is bio assays. These can be formed in to homogeneous groups of required size by equalizing them in respect of age, breed, housing and other management, etc. to form blocks. If the number of doses is small a latin square design can also be adopted to increase further the precision of estimate of the relative potency.

It is observed that in a 4 point assay a test of linearity of regression is

not available, as the deviation from the regression contrasts does not exist in the analysis of variance. In a 6 point assay there are 2 degree of freedom for deviation from regression. The following two contrasts of the dose totals are the two deviation contrasts.

Quadratic (L_2) combined = $(S_1 - 2S_2 + S_3) + (T_1 - 2T_2 + T_3)$

Difference between quadratic = $(S_1 - 2S_2 + S_3) + (T_1 - 2T_2 + T_3)$

In general the contrast L_n is represented by the n-th degree polynomial. The difference between two such polynomials is denoted as L_n'.

Incomplete Block Designs for Bio-Assays

When the number of doses is large it may not always be possible to get suitable homogeneous groups of experimental units for adopting randomized block designs. If litters are used as blocks a sufficient numbers of litters of required size may not be available even if the number of doses is not very large. This forces one to use incomplete block designs.

The use of incomplete blocks design in bio assays is limited due to inflexibility of existing incomplete block designs including balanced incomplete block-(BIB) designs. These designs aim at estimating all paired treatment differences, while in bio assays all contents are not of equal importance. In symmetrical parralled line assays and slope ratio assay only two major type of contrasts are of interest. The contrasts are :

i) The difference between the totals of standard and test preparations (the preparation contrasts L_P)
ii) The pooled estimate of slope (combined regression contrast L_1)

By designing properly it is possible to estimate the preparation and regression contrasts free from the block effects. In such designs the two contrast L_P and L_1 can be estimated by the unadjusted dose totals as in randomized block designs. Let $s_1, s_2 \ldots\ldots\ldots\ldots s_k$ denote K doses (on the log scale) of standard preparation arranged in ascending order and $t_1 \ldots\ldots\ldots, t_k$ that of test preparation are arranged is descending order of magnitude, the constant difference for successive log doses is the same for both the parameters. Let S_i and T_j denote the response totals corresponding to the doses S_i and T_j (i, j = 1, 2 .....k) The important contrast are (1) preparation contrast :

$$L_P = \sum_{i=1}^{k} (T_i - S_i)$$

The next contrast is combined regression contrast L_1 which is the

sum of the linear contrasts of dose totals of the two preparations. The two contrasts L_p and L_1 are essential for estimating the relative potency. The other contrasts provide validity tests and are known as parrallelism contrast and are denoted by L_1' (difference between two linear contrasts) L_2 and L_2' (sum and difference of the quadratic contrasts of the two doses totals) and pair of the contrasts of the type Lm and Lm' (sums and difference of the m^{th} power contrasts among the dose totals of the two preparations).

When an incomplete block design is used, the blocks are no longer orthogonal to the doses and at least some of the defined contrasts is not free from the block effects. Dose effects must therefore, be estimated by the method appropriate to the design and used in place of the unadjusted dose totals to calculate the contrast.

Some series of incomplete block designs for symmetrical parallel line assay from which the preparation contrast L_p and combined regression contrast L_1 can be obtained from the unadjusted dose totals as in the case of complete block designs. Take an incomplete block design for K doses of standard preparation in block size k′ (k′ < K). The final design is obtained by augmenting each block with k′ of test preparation with the rule that if s_1 doses of standard preparation is in the block then t_1 of the test preparation must be included in the block. Thus we have an incomplete block design for bio assays in 2k′ units per block from which both preparation and combined regression contrasts can be obtained from the unadjusted totals. The preparation and combined regression contrasts can be obtained from the unadjusted totals. The preparation contrast is independent of the block effect because equal number of doses of both preparations is occuring in the same block of the design. Since the dose totals S_1 and T_1 have the same coefficient but with the opposite sign, in the combined regression contrast, this contrast can be expressed as a weighted totals of within block contrast and is therefore, free from the block effects. For the same reason the contrasts L_{2n+1} and L^1_{2n+2} are estimated free from the block effects. The other constants L^1_{2n+1} and L_{2n+2} are not free from the block effects and therefore, must be estimated by the methods appropriate to the design. If is original design is BIB design then all the estimated contrasts are uncorrelated.

Example : An 8-point parallel line assay.

Take the BIBD with K=4, k′ =2 and use s_1, s_2, s_3 and s_4 as treatment.

Blocks		
	1	s_1 s_2
	2	s_1 s_3
	3	s_1 s_4

4 s_2 s_3
5 s_2 s_4
6 s_3 s_4

The final design in four plot block along with the doses of the test preparation is :

Blocks	1	s_1 s_2	t_1 t_2
	2	s_1 s_3	t_1 t_3
	3	s_1 s_4	t_1 t_4
	4	s_2 s_3	t_2 t_3
	5	s_2 s_4	t_2 t_4
	6	s_3 s_4	t_3 t_4

The coefficient of the dose totals in the different contrasts of the 8 point assay are

Table : Coefficient to the dose totals in the different contrasts of an 8 point Assays

Doses		**Standard preparation**				**Test Preparation**			
		s_1	s_2	s_3	s_4	t_1	t_2	t_3	t_4
Dose Totals		S_1	S_2	S_3	S_4	T_1	T_2	T_3	T_4
Con-	Lp	-1	-1	-1	-1	1	1	1	1
trasts	L_1	-3	-1	1	3	3	1	-1	-3
	L_1'	-3	-1	1	3	-3	-1	1	3
	L_2	1	-1	-1	1	1	-1	-1	1
	L_2'	1	-1	-1	1	-1	1	1	-1
	L_3	1	-3	3	-1	-1	3	-3	1
	L_3'	1	-3	3	-1	1	-3	3	-1

Except Lp, other contrast are orthogonal Tables (Fisher and Yates (1963)).

The estimate of regression contrasts from the dose totals is :

$$L_1 = -3(S_1-T_1) - (S_2-T_2) + (S_3-T_3) + 3(S_4-T_4).$$

Here $S_i - T_i$ ($i = 1, 2, 3, 4$) occur in the same block, $S_i - T_i$ is free from block effect. Lp which is a function of $S_i - T_i$ is also estimated free from

the block effect. From the same reason L_2' and L_2 are also free from the block effects. Here all contracts L^1_{2n} and L^2_{2n-1} for different values of n are estimated free from the block effects. The other contrasts. L_1', L_2 and L_3' are not free from blocks. To estimate them the individual dose effects are first estimated. These estimated dose effects are then substituted for the unadjusted dose totals in the affected contrasts.

Their sum of squares is found by squaring each of them and then dividing by the coefficient of σ^2 in the variance of the contrasts of the dose effects. If the initial design is BIBD the estimate of the affected contrasts obtained as indicated above are mutually orthogonal. By adding the sums of squares due to all these contrasts including the unaffected ones, the adjusted dose sum of squares is obtained. The error sum of square can now be obtained by subtraction.

In case of other incomplete block designs, the estimates of the affected contrasts may not he orthogonal. In these error sum of squares has to be obtained from the complete analysis of the final incomplete designs adopted for the assay.

Analysis of Parallel Line Assays Based on BIB Designs

The analysis of parallel line assays is essentially in two parts. The first part consists of computation of the analysis of variance and the second part consists of estimating the relative potency, its variance, and limits.

The expression for calculating sum of square due to any contrast unaffected by block effect is

$$\left[\sum_{i=1}^{k} l_i S_i - \sum_{i=1}^{k} l_i T_i\right]^2 \Big/ r\left(2\sum_{i=1}^{k} l_i^2\right)$$

Where l_1, l_2 ..... the coefficient of the contrasts and r is the number of replications. The sum of square due to other contrasts are calculated from least squares estimates s_1, s_2,..... s_k and t_1 t_2,..... t_k of the effects of various doses of the standard and test preparations (we use the same symbols for doses and their effects as there is no confusion). The details of analysis for designs based on BIB design is given below.

The adjusted totals are

$$Q_i^s = S_i - \sum_{j(i)} \bar{y}_j$$

$$Q_i^t = T_i - \sum_{j(i)} \bar{y}_j$$

Where $\bar{y}_j$ (j = 1, 2,....., b) is the *jth* block average and summation is taken over all the block containing dose i of the preparations. All the required contrasts are functions Q's. The unaffected contrasts are functions of $Q^s_i - Q^t_i (= S_i - T_i)$; the sum of squares for this are conveniently calculated from unadjusted dose totals. The remaining contrasts are function of

$$Q_i = Q_i^s + Q_i^t$$

The s.s. due to any affected contrast $\left[\sum_{i=1}^{k} l_i Q_i\right]$ is given by

$$\left[\sum_{i=1}^{k} l_i Q_i\right]^2 \Big/ rE\left(\sum_{i=1}^{k} l_i^2\right)$$

where E = (rk' - r + λ)/rk', k' being the block size of the starting design and λ the number of blocks in each of which any two given doses of standard preparation occur together.

The variance of the affected contrast is $\left(2\sum_i l_i^2 / rE\right)\sigma^2_{2k}$, where σ^2_{2k} is the error variance. If a randomized block design is used then the corresponding variance is $\left(2\sum_i l_i^2 / r\right)\sigma^2_{2k}$ where σ^2_{2k} denotes the error variance in complete blocks with *2k* units. If one uses a BIB design with *2k* doses, block size K_1 then the variance $\left(2\sum_i l_i^2 / rE'\right)\sigma^2_{2k}$ where *E'* is the efficiency factor of BIB design and σ^2_{2k} is its error variance. The

efficiency of the present design in respect of affected contrast is $E\,\sigma^2_{2k} / \sigma^2_{2k'}$ when compared to randomized blocks design and $(E/E')\left(\sigma^2_{k_i} / \sigma^2_{2k'}\right)$, when compared with the basic BIB design.

An Example : For illustrating the method of analysis of a 6 point symmetrical parallel line assay Bliss (1952. p. 498) presented a body of data collected on a vitamin D assay, the design used *12* litter of 6 rats as randomized blocks. Here we used these data with some modifications. To ensure comparability of the estimate of relative potency all observation were used, but were fitted into an incomplete block design of the present series by omitting two observations from each of the original block (litters), as shown by blanks in table below, and forming 6 additional block (13-18) from the 24 observations omitted, ignoring litter differences, but retaining the dose-observation relations.

Table : Data and BIB based on Parallel Line Assay (Doses in mg.)

Block	Standard			Test			Block
Number	S_1	S_2	S_3	t_1	t_2	t_3	Total
	2.5	5	10	2.5	5	10	
1	2	8	-	-	9	7	26
2	6	-	9	3	-	8	26
3	-	6	12	4	6	-	28
4	9	11	-	-	14	13	47
5	10	-	17	8	-	10	45
6	-	7	5	6	9	-	27
7	4	10	-	-	11	13	38
8	11	-	9	3	-	15	38
9	-	9	14	5	8	-	36
10	4	7	-	-	10	10	31
11	12	-	9	15	-	15	51
12	-	8	11	7	8	-	34
13	4	4	-	-	5	9	22

14	7	-	8	3	-	9	27
15	-	15	10	6	8	-	39
16	2	4	-	-	6	6	18
17	4	-	13	5	-	12	34
18	-	10	13	4	18	-	45
Totals	75	99	130	69	112	127	612
	S_1	S_2	S_3	T_1	T_2	T_3	
Adjusted	-25.75	1.25	22.50	-31.75	14.25	19.50	0
Total	Q_1^S	Q_2^S	Q_3^S	Q_1^T	Q_2^T	Q_3^T	

S.S. due to block = 358.00 contrast for 6 points parrallel line assay.

Table : Contrast, Divisor and Sums of Squares for Data of Table

	Value of Contrast	Divisor for SS	Sum of Squares
Unaffected contracts			
Lp	4	6 x 12 = 72	0.22
L_1	113	4 x 12 = 48	266.02
L'_2	-35	12 x 12 = 144	8.51
Affected Contrasts			
L_1	-3.00	4 x 9 = 36	0.25
L_2	-46.50	12 x 9 = 108	20.02
Total			295.02

For this design, $r = 12$, $k' = 2$ and $\lambda = 6$, whence $E = (rk' - r + \lambda) / k' = 9$. The analysis of variance is in the following table.

As in the original analysis, all the validity tests are satisfied. Though the estimate of relative potency must remain identical with the original estimate, the mean squares due to error and the affected contrast will be different. After regrouping the data there are 49 *d.f.* for error, of which *10 d.f.* may be contaminated by differences between the original litters. The

mean square for the contaminated component is 6.11 while that for the remaining 39 d.f. is 7.23 : the error mean square in the original analysis was 7.22 with 55 d.f. Thus grouping of observations on animals from different litters into the same block has not increased the error mean square (as might have been expected from the significance of the between block means square). The estimate of the relative potency is

$$R = \frac{2.5}{2.5}\text{antilog}\left\{\frac{4d}{3}.\frac{L_P}{L_1}\right\} = \text{antilog}\left\{\frac{16d}{339}\right\}$$

where $d = \log_{10}2$

Table : Analysis of Variance Table

Nature of variation	d.f.	s.s.	m.s.	F
Between Blocks	17	358.00	21.06	3.01 **
Preparations	1	0.22	0.22	
Regression	1	266.02	266.02	38.00 **
Parrallelism	1	0.25	0.25	< 1
Comb. Quadratic	1	20.02	20.02	2.86 NS
Diff. Quadratic	1	8.51	8.51	1.22 NS
All Contrasts	5	295.02		
Error (by subtraction)	49	342.98	7.00	
Total	**71**	**996.00**		

❑❑❑

19

EXPECTED VALUE OF MEAN SQUARE IN FIXED, RANDOM AND MIXED MODELS

At the planning stage the experimenter must decide whether the levels of factors are to he taken as fixed or are to be chosen at random from many possible levels. This decision must be made prior to the running of the experiment and if random levels are to be used, they must be chosen from all possible levels by a random process. In the case of random levels it will be assumed that the levels are chosen from an infinite population of possible levels. When all the levels are fixed, the mathematical model of the experiment is called a fixed model. When all levels are chosen at random, the model is called a random model. When several factors are involved, some at fixed levels and others at random levels, the model is called a mixed model.

Single factor experiment

If the design is completely randomized, the model is $x_{ij} = \mu + T_j + e_{ij}$

Whether the treatment levels are fixed or random. It is assumed that μ is a fixed constant and the errors are normally and independently distributed with zero mean and same variance $e_{ij} \sim N(0, \sigma_e^2)$. The assumption under fixed and random model will be.

Fixed model

1. Assumption :
   T_j's are fixed constants

$$\sum_{j=1}^{k} T_j = \sum_{j=1}^{k} (\mu_{.j} - \mu) = 0$$

Random Model

1. Assumption : T_j's are random variables with mean zero and constant variance σ_T^2.

(These add to zero as they are the only treatment means being considered,

(Here σ_T^2 represents the variance among $T_J^{'s}$ or among the true treatment means $\mu_j{}^{'s}$. The $T_J^{'s}$ average to zero when averaged over all possible levels, but for the K levels of the experiment they usually will not average to 0)

2. Analysis : Procedure is the same as given in earlier chapter for computation of SS

2. Analysis : same as for fixed model

3. EMS (Expected value of mean square)

Source	d.f.	EMS
Treatment	$k-1$	$\sigma_e^2 + \frac{\sum_{j=1}^{k} T_j^2}{k-1}$
Error	k(n-1)	σ_e^2

3. EMS (Expected value of mean square)

Source	d.f.	EMS
Treatment	$k-1$	$\sigma_e^2 + n\sigma_T^2$
Error	k(n-1)	σ_e^2

4. Hypothesis tested

$H_o : T_j = 0$ for all j

4. Hypothesis tested

$H_o : \sigma_T^2 = 0$

The EMS (expected mean square) column is extremely important in deciding how to set upon F test for significance. For the fixed model if the hypothesis is true that is $T_J = 0$ for all j *i.e.* all the fixed treatment means are equal then $\Sigma T^2_j = 0$ and EMS for T_j and error, are both $\sigma^2 e$.

Hence, the observed mean squares for treatment and error mean square are both estimates of error variance and they can be compared by means of a F-test. If this F test shows a significant high values, it must mean that $\frac{n\Sigma T_j^2}{k-1}$ is not zero and the hypothesis is to be rejected.

The expected mean square for any term in the model is the long range average of the calculated mean square when the x_{ij} from the model

is substituted in algebraic form in to the mean square computation.

For the random model, if the hypothesis is true, $\sigma_T^2 = 0$ *i.e.* variance of all treatment means is zero, then again each mean square is an estimate of error variance. Again an F-test between the two mean square is appropriate.

So it is seen that for a single factor experiment there is no difference in the test to be made after the analysis and the only difference is in the generality of the conclusions. If Ho is rejected, there is probably a difference between the K fixed treatment means for the fixed model; for the random model there is a difference between all the treatments of which the K examined are but a random sample. The EMS for treatment effect is often written as $\sigma_e^2 + n\sigma_t^2$ whether the model is fixed or random.

Two factor model :

For two factors A and B the model is in case of RBD. $X_{ijk} = \mu + A_i + B_j + (AB)_{ij} + e_{k(ij)}$ i =1,2 ..... a, j = 1, 2 ..... b k = 1, 2.....n.

Again μ is a fixed constant and $e_{k(ij)}$ are normally and identically distributed with mean O and variance , σ_e^2 . If both A and B are at fixed levels, the model is a fixed model. If both are at random level the model is a random model, and if one is at fixed levels and the other at random levels, the model is a mixed model. comparising of each of these models gives.

	Fixed	Random	Mixed
1.	Assumption	Assumption	Assumption
	A_i's are fixed constants	A_i's are NID (o, σ_A^2)	A_i's are fixed
	$\sum_{i=1}^{a} A_i = 0$		$\sum_{i=1}^{a} A_i = 0$
	B_j's are fixed Constants and	B_j's are NID (o σ_B^2)	B_j are NID $(0, \sigma_B^2)$
	$\sum_{j=1}^{b} B_j = 0$		
	$(AB)_{ij}$ are fixed	$(AB)_{ij}$ are NID (o, σ_{AB}^2)	$(AB)_{ij}$ are NID

Constants and		(o, σ^2_{AB}) and
$\sum_{i=1}^{a}\sum_{j=1}^{b}(AB)_{ij} = 0$		$\sum_{i=1}^{a}(AB)_{ij} = 0$
		$\sum_{j=1}^{b}(AB)_{ij} \neq 0$
		(for A fixed and B random)
2. Analysis Procedure is same as described earlier for sums of square	Same	Same

Source	d.f	EMS (fixed)	EMS (Random)	EMS (Mixed)
A_i	a-1	$\sigma_e^2 + nb\frac{\sum_{i=1}^{a} A_i^2}{a-1}$	$\sigma_e^2 + n\sigma_{AB}^2 + nb\sigma_A^2$	$\sigma_e^2 + n\sigma_{AB}^2 + nb\frac{\sum_{i=1}^{a} A_i^2}{a-1}$
B_j	b-1	$\sigma_e^2 + na\frac{\sum_{j=1}^{b} B_j^2}{b-1}$	$\sigma_e^2 + n\sigma_{AB}^2 + na\sigma_B^2$	$\sigma_e^2 + n\sigma_B^2$
$(AB)_{ij}$	(a-1) (b-1)	$\sigma_e^2 + \frac{n\sum_i\sum_j AB_{ij}^2}{(a-1)(b-1)}$	$\sigma_e^2 + n\sigma_{AB}^2$	$\sigma_e^2 + n\sigma_{AB}^2$
$e_{k(i,j)}$	ab (n-1)	σ_e^2	σ_e^2	σ_e^2

4. Hypothesis tested

$H_1 : A_i = 0$ for all i	$H_1 : \sigma^2 A = 0$	$H_1 : Ai = 0$ for all i
$H_2 : B_j = 0$ for all j	$H_2 : \sigma^2 B = 0$	$H_2 : \sigma_B^2 = 0$
$H_3 : AB_{ij} = 0$ for all i and j	$H_2 : \sigma^2 AB = 0$	$H_3 : \sigma_{AB}^2 = 0$

Summing the interaction term over the fixed factors (Σi)is zero but

summing it over the random factor (Σj) is not zero affects the expected mean squares.

For the fixed model, the mean square for A, B and AB are each compared to the error mean square to test the respective hypothesis. For the random model the third hypothesis of no interaction is tested by comparing the mean square for interaction to the mean square for error. But the first, second hypothesis are each tested by comparing the mean square for main effects (Ai or Bi) with the mean square for the interaction as can be seen by their expected mean square values. For the mixed model, the interaction hypothesis is tested by comparing the interaction, mean square by the error mean square. The random effect (Bj) is also tested by comparing its mean square with the error mean square. The fixed effect (Ai), are however tested by comparing its mean square with the interaction mean square.

So it can be seen that how important is the EMS column, it tells how the tests of hypothesis will be done. So it is important that these EMS expression are determined prior to the running of the experiment. In some cases the proper test indicated by EMS column will show insufficient degree of freedom and where the investigators might like to change the experiment.

EMS rules

Because of the importance of EMS column in determining what tests of significance are to be done after the analysis is completed. In complex models this EMS column is again very important and to find the EMS value it is very difficult so it is useful to have some simple method of determining the EMS values from the models of the given experiment (Hicks, 1964). A set of rules will determines the EMS column very rapidly without recourse to their derivation. To determine the EMS column for any model: the rules will be illustrated for the two factor mixed model as given in Hicks(1964).

1. Write the variable terms in the model as row heading in a two way table

A_i
B_j
$(AB)_{ij}$
$e_{k(ij)}$

2. Write the subscripts in the model as column heading : over each

subscript write F if the factor levels are fixed, R if random; also write the number of observation each subscript is to cover.

	a	b	n
	F	R	R
	i	j	k
A_i			
B_j			
$(AB)_{ij}$			
$e_{k(ij)}$			

3. For each row (each term in the model) copy the number of observations under each subscript, providing the subscript does not appear in the row heading.

	a	b	n
	F	R	R
	i	j	k
A_i		b	n
B_j	a		n
$(AB)_{ij}$			n
$e_{k(ij)}$			

4. For any bracketed subscript in the model place a1 under those subscripts which are inside the brackets.

	a	b	n
	F	R	R
	i	j	k
Ai		b	n
Bj	a		n
$(AB)_{ij}$			n
$e_{k(ij)}$	1	1	

5. Fill the remaining cells with 0 or 1, depending upon whether the subscript represents a fixed (F) or a random (R) factors.

	a	b	n
	F	R	R
	i	j	k
A_i	o	b	n
Bj	a	I	n
$(AB)_{ij}$	o	I	n
$e_{k(ij)}$	I	I	I

6. To find the expected mean square for any term in the model. Cover the entries in the column (or columns) which contain non bracketed subscripted letters in this term in the model (*e.g.* for Ai, cover column i; for $e_{k(ij)}$ cover column k)

Multiply the remaining numbers in each row. Each of these products is the coefficient to its corresponding term in the model provided the subscript on the term is also a subscript on the term whose expected mean square is being determined. The sum of these coefficient multiplied by the variance of their corresponding terms is the expected mean square of the term being considered (*e.g.*, for Ai, cover colùmn i). The products of the remaining coefficients are bn, n, n and I, but the term n is not used, as there is no i in its term (Bj). The resulting expected mean square is then (bn σ^2A + nσ^2AB + 1.σ^2e). For all terms these rules give:

	a F i	b R j	n R k	EMS
A_i	o	b	n	$\sigma_e^2 + n\sigma_{AB}^2 + nb\sigma_A^2$
B_j	a	I	n	$\sigma_e^2 + na\sigma_B^2$
$(AB)_{ij}$	o	I	n	$\sigma_e^2 + n\sigma_{AB}^2$
$e_{k(ij)}$	I	I	I	σ_e^2

Here σ_A^2 is a fixed type of variance $\sigma_A^2 = \frac{\sum_{i=1}^{a} A_i^2}{a-1}$

The application of these rules becomes very easy to use after a bit of practice.

Example :

Consider a 6 x 4 x 5 x 2 factorial experiment. Here the first factor Oi has random levels (6 being chosen). Factor A_j has 4 fixed levels, C_k has 5 fixed levels and (L_M) has two fixed levels. There were six replications and the design used is CRD. The expected mean square can be determined from the rules given in earlier section.

EMS values

	6	4	5	2	6	
	R	F	F	F	R	
Model	i	j	k	m	q	
O_i	1	4	5	2	6	$\sigma_e^2 + 240\ \sigma_0^2$
A_j	6	0	5	2	6	$\sigma_e^2 + 60\ \sigma_{0A}^2 + 360\ \sigma_A^2$
OA_{ij}	1	0	5	2	6	$\sigma_e^2 + 60\ \sigma_{0A}^2$
C_k	6	4	0	2	6	$\sigma_e^2 + 48\ \sigma_{0C}^2 + 288\ \sigma_C^2$
OC_{ik}	1	4	0	2	6	$\sigma_e^2 + 48\ \sigma_{0C}^2$
AC_{jk}	6	0	0	2	6	$\sigma_e^2 + 12\ \sigma_{0AC}^2 + 72\ \sigma_{AC}^2$
OAC_{ijk}	1	0	0	2	6	$\sigma_e^2 + 12\ \sigma_{0AC}^2$
L_m	6	4	5	0	6	$\sigma_e^2 + 120\ \sigma_{0L}^2 + 720\ \sigma_L^2$
$D\,L_{im}$	1	4	5	0	6	$\sigma_e^2 + 120\ \sigma_{0L}^2$
$A\,L_{jm}$	6	0	5	0	6	$\sigma_e^2 + 30\ \sigma_{0AL}^2 + 180\ \sigma_{AL}^2$
OAL_{ijm}	1	0	5	0	6	$\sigma_e^2 + 30\ \sigma_{AL}^2$
CL_{km}	6	4	0	0	6	$\sigma_e^2 + 24\ \sigma_{0CL}^2 + 144\ \sigma_{CL}^2$
OCL_{ikm}	1	4	0	0	6	$\sigma_e^2 + 24\ \sigma_{0CL}^2$
ACL_{jkm}	6	0	0	0	6	$\sigma_e^2 + 6\ \sigma_{0ACL}^2 + 36\ \sigma_{ACL}^2$
$OACL_{ijkm}$	1	0	0	0	6	$\sigma_e^2 + 6\ \sigma_{0ACL}^2$
Eq (i, j, k, m)	1	1	1	1	1	σ_e^2

Here it is easily seen that all interaction involving factor O and main effect O are tested against the error mean square at the bottom of the table. All interaction and the main effects involving fixed factors are tested by the mean square just below them in the table. However algebraic derivations of EMS is also shown as below.

EMS Derivations :

Single factor experiment. The model is

$$X_{ij} = \mu + T_j + e_{ij} \qquad \text{(i)}$$

Here the sum of squares for the treatment effect is given as :

$$SST = \sum_{j=1}^{k} n(\bar{x}.j - \bar{x}..)^2$$

Here the number of observation per cell per treatment is equal to n.

$$\bar{x}.j = \sum_{i=1}^{n} x_{ij} / n = \sum_{i=1}^{n} \frac{(\mu + T_j + e_{ij})}{n}$$

$$\bar{x}.j = \frac{n\mu}{n} + \frac{nT_j}{n} + \sum_{i=1}^{n} e_{ij} / n$$

$$\bar{x}.j = \mu + T_j + \sum_{i=1}^{n} e_{ij} / n$$

Also $$\bar{x}.. = \sum_{j=1}^{k}\sum_{i=1}^{n} x_{ij} / nk = \sum_{j}^{k}\sum_{i}^{n}(\mu + T_i + e_{ij}) / nk$$

$$\bar{x}.. = \frac{nk\mu}{nk} + \left(n\sum_{j=1}^{k} T_j / nk\right) + \left(\sum_{j=1}^{k}\sum_{i=1}^{k} e_{ij} / nk\right)$$

$$= \left(\mu + \sum_{j=1}^{k} T_j / k + \sum_{j=1}^{k}\sum_{i=1}^{n} e_{ij} / nk\right)$$

we have

$$\bar{x}.j - \bar{x}.. = T_j - \left(\sum_{j=1}^{k} T_j / k\right) + \left(\sum_{i=1}^{n} e_{ij} / n\right) - \left(\sum_{i=1}^{n}\sum_{j=1}^{k} e_{ij} / nk\right)$$

squaring gives

$$(\bar{x}.j - \bar{x}..)^2 = \left[T_j - \left(\sum_{j=1}^{k} T_j / k\right)\right]^2 + \frac{1}{n^2}\left[\sum_{i=1}^{n} e_{ij} - \sum_{i}^{n}\sum_{j}^{k} e_{ij} / k\right]^2$$

+ crossproduct term.

Multiplying by n and summing over j gives.

$$SST = \sum_{j=1}^{n} n(\bar{x}.j - \bar{x}..)^2 = n\sum_{j=1}^{k}\left[T_j - \sum_{j=1}^{k} T_j / k\right] + \frac{n}{n^2}$$

$$\sum_{j}\left[\sum_{i=1}^{n} e_{ij} - \left(\sum_{i}^{n}\sum_{j}^{k} e_{ij} / k\right)\right]^2 + n\sum_{j}(\text{crossproduct})$$

$$E(SST) = nE\left[\sum_{j=1}^{n}\left(T_i - \sum_{j=1}^{k} T_i / k\right)^2\right] + \frac{1}{n}$$

$$E\left\{\sum_{j=1}^{k}\left[\sum_{i=1}^{n} e_{ij} - \left(\sum_{i=1}^{n}\sum_{j=1}^{n} e_{ij} / k\right)\right]^2\right\}$$

as it can be shown that expacted values of crossproduct term is equal to zero.

It the treatment levels are fixed.

$$\sum_{j=1}^{k} T_j = 0$$

and E(SST) become

$$E(SST) = n\sum_{j=1}^{k} T_j^2 + \frac{1}{n}(nk - n)\sigma_e^2$$

Error component are random and $\sum_{j=1}^{k} T_j^2$ is a constant. The E(SST/ k-1) so

$$E(MST) = \left(n\sum_{j=1}^{k} T_j^2 / k - 1\right) + \frac{n(k-1)}{n(k-1)}\sigma_e^2$$

If however the treatment levels as random.

$$\sum_{j=1}^{k} T_j \neq 0$$

$$E\left[\Sigma(x_i - \bar{x})^2\right] = (n-1)\sigma_x^2$$

$$E(MST) = \frac{n(k-1)\sigma_T^2}{(k-1)} + \sigma_e^2$$

For the error mean squre

$$SSE = \sum_{j=1}^{k}\sum_{i=1}^{n}\left(x_{ij} - \bar{x}.j\right)^2$$

we have

$$x_{ij} - \bar{x}.j = e_{ij} - \sum_{i=1}^{n} e_{ij} / n$$

Squaring and adding

$$\sum_{j=1}^{k}\sum_{i=1}^{n}\left(x_{ij} - \bar{x}.j\right)^2 = \sum_{j=1}^{k}\sum_{i=1}^{n}\left[e_{ij} - \left(\Sigma e_{ij} / n\right)\right]^2$$

Taking the expected value, we have.

$$E(SSE) = E\sum_{j=1}^{k}\sum_{i=1}^{n}\left[e_{ij} - \left(\Sigma e_{ij} / n\right)\right]^2$$

$$= \sum_{j=1}^{k} E \sum_{i}\left[e_{ij} - \left(\Sigma e_{ij} / n\right)\right]^2$$

$$= \sum_{j=1}^{k}(n-1)\sigma_e^2 = k(n-1)\sigma_e^2$$

$$E(MSE) = E\left(\frac{SSE}{K(n-1)}\right) = \sigma_e^2$$

So the expected values are coming out to the same as we have shown in table using simple rules. So there is a advantage of simple rules in determining the EMS values.

❑❑❑

References

1. Bliss C.I. (1952). The Statistics of Bio-Assay with special Reference to Vitamins Vol. II Academic Press, New York.
2. Cochran, W.G and Cox, G.M. (1966) Experimental Designs, Wiley, New York.
3. Das M.N. and Giri N.C. (1991) Design and Analysis of Experiments. Wiley Eastern, New Delhi.
4. Dey A. (1986). Theory of Block Designs, Wiley Eastern, New Delhi.
5. Federer, W.T. (1955) Experimental Design Theory and Application, Macmillian and Company, London.
6. Finney, D.J. (1978). Statistical Methods in Biological Assay 3rd edition Griffine, London.
7. Gomez, Kwanchai A. and Gomez, Arturo, A (1984). Statistical Procedures for Agricultural Research 2nd edition. An International Rice Research Institute Book. John Wiley and Sons, New York.
8. Gupta S.C. and Kapoor, VK (1983) Fundamentals of Applied Statistics, Sultan Chand & Sons, N. Delhi.
9. John PWM; (1971) Statistical Design and Analysis of Experiments. Collier-Mecmillan, London.
10. Hicks, Charles R. (1964) Fundamental Concepts in the Design of Experiments, Holts, Rinehart and Winston, New York.
11. Kempthorne, O (1991) Design and Analysis of Experiments, John Wiley, New York.
12. Narain, Prem, Bhatia VK and Malhotra, PK (1979) Hand Book of Statistical Genetics IASRI, New Delhi.

13. Nigam AK, Gupta, VK (1979) Hand Book on Analysis of Agricultural Experiments. IASRI, New Delhi.

14. Nigam, AK, Puri PD and Gupta, VK (1988) Characterization and Analysis of Block Designs. Wiley Eastern Limited, New Delhi.

15. Prasad, Rajainder and Jaggi, Seema (2003) Efficient Designing of Experiments and Analysis of Experimental data. Reference Manual volume I & II, IASRI, New Delhi.

16. Raghavarao, D (1971) Construction and Combinational Problems in Design of Experiments, Wiley, New York.

17. Rai Lajpat and Grover, Deepak (2005) Manual on Designing and Analysis of Agricultural data. Deptt. of Mathematics & Statistics, College of Basic Scs & Humanities, CCSHAU, Hisar.

18. Rao, G. Nageshwara (1983) Statistics for Agricultural Sciences, Oxford and IBH Publishers, New Delhi.

19. Roger G.P. (1994) Agricultural Field Experiments, Design and Analysis, Marched Dekker, Inc. New York.

20. Singh, RK and Caudhary, BD (1979), Biometrical Methods in Quantitative Genetic Analysis. Kalyani Publishers, New Delhi.

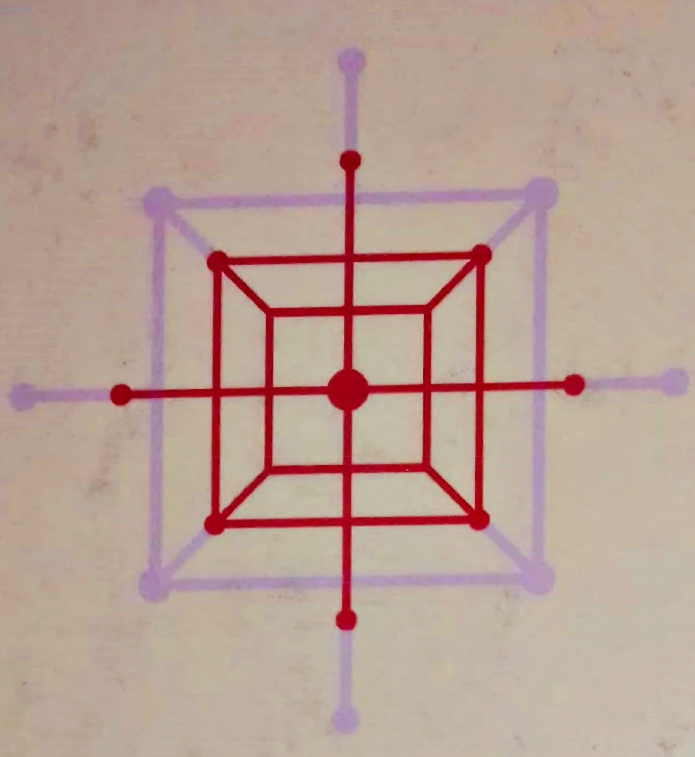
21. Srivastava, R. and Prasad, Rajinder (2000). Design and Analysis of Agricultural Experiments, Centre of Advance Studies on Agricultural Statistics and Computer Applications, IASRI, New Delhi.

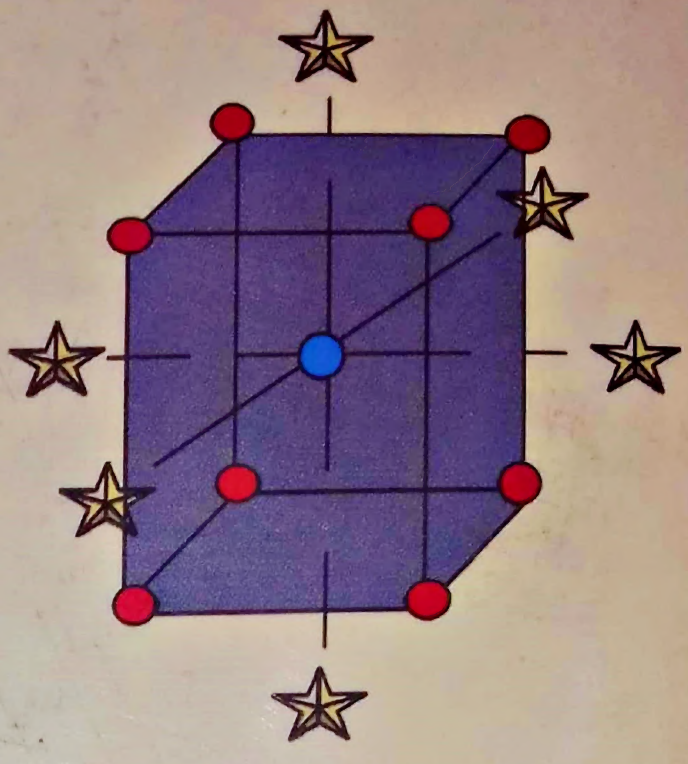
❐ ❐ ❐

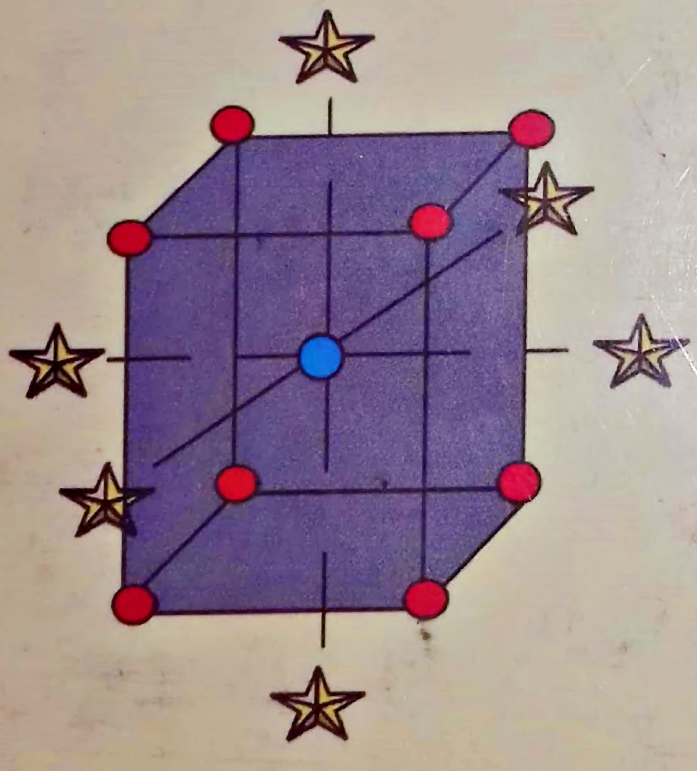
SUBJECT INDEX

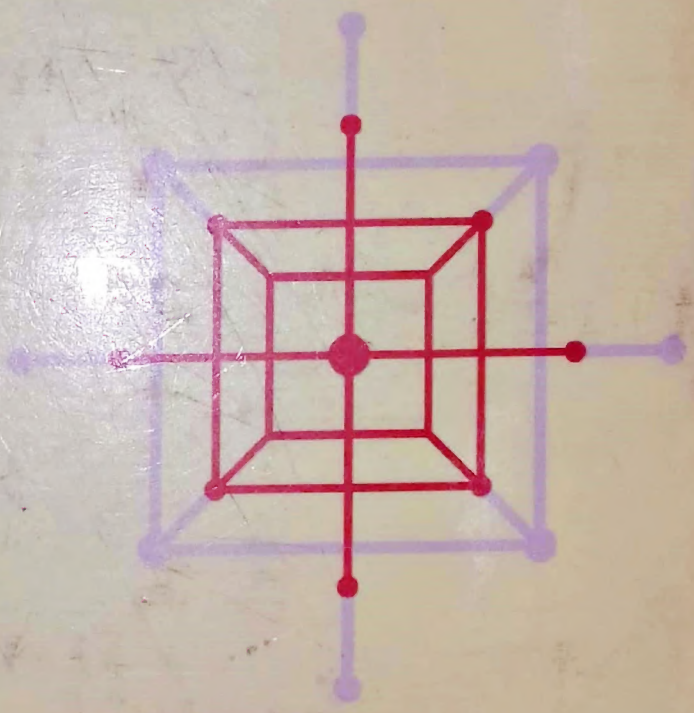
absolute experiment 20
adjusted treatment..... 55,106
advantages of confounding ... 135
advantages of RBD 64
agronomy trails 23
analysis of variance 15
angular transformation 41
anova table 27
A-optimum designs 234
assumptions 172
augmented BIB 225
augmented CRD 214
augmented design 214
augmented GD designs 241
augmented LSD 214
augmented PBIB designs 241
augmented RBD 214
balance incomplete design ... 191
balanced design 146
balanced lattice 202
Basic principles 9,12
BIBD 191,229
biological assays 274
block designs 229
blocking 24
choice of design 18
combined analysis..... 186
combining ability..... 252
combining ability analysis 258
comfounding 134
comparative experiment..... 20
complete confounding..... 135
controlled condition 10
correction factor 48
covariance analysis 172
CRD 45
critical value..... 13
cross over designs..... 75
data analysis 160
degree of freedom 14
degree of precision 12
demerits of CRD 47
design of experiments 9,12
dialled crosses 248
disadvantages of confounding 135
disadvantages of RBD 66
D-optimum designs 234
draw conclusions 188
Duncan's multiple test..... 29
Dunnett's test..... 28
efficiency 195
efficiency of a design 22
elementary contrasts..... 244
EMS derivations 297
EMS rules 293
E-optimum design 235
error mean square 14
expected value of mean 289
experimental errors 21,180
experimental material 17
experimental unit 21
factorial experiment 23,110
factors..... 22
field plan 219

fixed effect model ... 257,263,269
fixed effects model 26
F-test 14
full diallel 252
GCA effects 260,265
GD design 235
genetic component 256
genetic resources 213
genotypic differences 264
Graeco LSD 68
Griffing's approach 248
hypothesis 17
IBD for bioassays 284
inter block information 196
interpretation of results 218
Iterative method 100
L_2 PBIB design 238
latin square design 16,60,63
lattice design 202
layout 46,53
layout and analysis 213
least significant difference 27
local control 16
main designs 19
mean performance 224
merits of CRD 46
metametric transformation 275
missing observation .. 96,107,155
missing plot technique 94
mixed models 289
multi-locational experiment .. 180
multiple comparision tests .....
..... 27,28,31
Neutman-Keuls test 28
number of replications 11,13
objective of the experiment 17
optimalety of designs 234
orthogonal contrasts 32
orthogonal latin squares 73
parallel line assays 276
partial confounding 137
partially balanced lattice 211
PBIBD 230
PEB design 231
planning consideration 16
plot size & shape 18
precision 22
probit transformation 42
random effect model .....
..... 36,256,263,269
random models 289
randomization 14,203
RBD 53,84
reciprocal effects 253
reciprocal transformation 41
replication 12,18
replication effect 39
right type of designs 11
salient features of RBD 53
sampling 82
SCA effects 266
Scheffe's test 31
simple rules 299
single factor experiments 45,289
specific combining ability 253
split plot design 88,148
split-split plot design 153
square root transformation 39









AGROTech
Rs. 200/-

Agrotech Publishing Academy
11-A, Vinayak Complex-B, Durga Nursary Road,
Udaipur - 313001 INDIA
Mobile : 9414169635